

Attach to Paper

Access DB# _____

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: Room 111 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

***For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.**

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr.Link _____
Date Completed: _____	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: _____	Other _____	Other (specify) _____

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2003, 15:52:40 ; Search time 323 Seconds

(without alignments)
6101.253 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426

Sequence: 1 tgtgtgtgtgtgtcactattt.....agaaagttatattccatca 6426

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

- 1: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
- 2: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
- 3: /cgn2_6/prodata/2/lna/6A.COMB.seq:*
- 4: /cgn2_6/prodata/2/lna/6B.COMB.seq:*
- 5: /cgn2_6/prodata/2/lna/PCRTUS.COMB.seq:*
- 6: /cgn2_6/prodata/2/lna/backfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	65.6	1.0	7218	1	US-08-232-463-14
2	58	0.9	240	1	US-08-628-417-6
3	56.4	0.9	2447	2	US-09-014-969-14
4	56	0.9	5361	4	US-08-973-462-2
5	56	0.9	6152	4	US-08-973-462-1
6	55.6	0.9	11485	4	US-09-410-464-9
7	54.6	0.8	6660	2	US-08-771-602D-1
8	54.6	0.8	6660	4	US-09-232-446B-1
9	54.4	0.8	3275	4	US-09-370-838-151
10	53.2	0.8	1798	4	US-09-797-906-1
11	53.2	0.8	8920	2	US-08-446-855A-1
12	53.2	0.8	8920	4	US-09-150-741-1
13	52	0.8	2223	4	US-08-257-073-4
14	52	0.8	30549	4	US-09-134-001C-322
15	51.4	0.8	1447	4	US-09-443-041A-27
16	50.8	0.8	3095	6	5231168-1
17	50.2	0.8	6671	1	US-08-280-443-1
18	50.2	0.8	6671	1	US-08-457-459-1
19	50.2	0.8	6671	1	US-08-555-678-1
20	50.2	0.8	6671	5	PCR-US95-02275-1
C 21	50	0.8	260	2	US-08-520-678A-29
22	50	0.8	260	4	US-08-897-126-29
23	50	0.8	1281	4	US-09-134-001C-509
24	49.6	0.8	5181	1	US-08-257-073-10
25	49.4	0.8	1117	4	US-09-247-373B-33
26	49.4	0.8	2674	4	US-09-817-180-1
27	48.6	0.8	2150	1	US-08-198-446B-10

28	48.6	0.8	2150	2	US-08-870-693-10	Sequence 10, Appl
29	48.6	0.8	2269	4	US-09-394-645-1	Sequence 1, Appl
30	48.6	0.8	2269	3	US-09-243-560B-1	Sequence 1, Appl
31	48.2	0.8	6755	3	US-08-931-999-4	Sequence 4, Appl
32	48	0.7	7101	2	US-08-480-604A-9	Sequence 9, Appl
33	48	0.7	7101	2	US-08-405-496A-9	Sequence 9, Appl
34	48	0.7	7101	4	US-08-915-136-9	Sequence 9, Appl
35	48	0.7	7101	4	US-08-957-310-9	Sequence 9, Appl
36	47.8	0.7	4766	5	PCT-US93-07261-10	Sequence 10, Appl
37	47.6	0.7	618	4	US-09-134-001C-1262	Sequence 10, Appl
38	47.4	0.7	140	1	US-08-628-417-5	Sequence 5, Appl
39	47.4	0.7	2187	4	US-09-134-001C-2131	Sequence 2131, Ap
40	47.2	0.7	454	2	US-08-623-906A-6	Sequence 6, Appl
41	47.2	0.7	2634	4	US-09-463-238-3	Sequence 3, Appl
42	47	0.7	1454	4	US-09-372-422A-19	Sequence 19, Appl
43	47	0.7	2186	4	US-09-360-545-66	Sequence 66, Appl
44	46.8	0.7	1051	4	US-09-245-041-10	Sequence 10, Appl
45	46.8	0.7	2082	2	US-08-785-310A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 130472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 1.0%, Score 65.6; DB 1; Length 7218;

Query Match	0.9%	Score 56	DB 4	Length 5361
Best Local Similarity	45.9%	Pred. No. 0.0036		
Matches 191	Conservative 0	Mismatches 225	Indels 0	Gaps 0
QY 3869	ACTGTTCCTAAAGTATATGAGACAGATGAAAAAATCTCTCCATTAATTCATTTAGATGATCAT	3928		
Db 289	AATGCTGAATTAGCTAGTAGGAAGAACTAAAGAAAAAATCTTGCACTTTTGAAGAAAGGA	348		
QY 3929	ACTGACATTGCTACGACTCGGGAAATGATVAGCAATTCACAGAAATCCGACATTCATCG	3988		
Db 349	AATCATTAACACTGAAAGCTGATGATTAATAAAATTTAGAAAGACCGAAGATTTAAAG	408		
QY 3989	AAAAATGAAATATATCCACGAGTGTATTAAGAAAAAATACTGAAATATATCCAAAACACATT	4048		
Db 409	GAAAAATATCTTAATTAAGTAATATAGAACAAACCAAAGAAAAAATTAATGACAAATTTATTA	468		
QY 4049	GAAGATATCTTCGTCGATAGAGAGATTGGATGTAATTTGGAACGTATATATGTCATGAATT	4108		
Db 469	AATATATATTGGACAAAAATTCAGAAAAAACAGAAATGTATCGAAAAATATACAAAGTCAGT	528		
QY 4109	GAGAAATGTGATTAATGACAGATGACATTGCTGAAGCTAAATCCATCCAGAGTGAATAAT	4168		
Db 529	GATGAACTTTTAAAGATTAATTAATATAGTGAAGATGTTAAATGAGAGATTAAGAAAAAAT	588		
QY 4169	GATGTCGATGAATGAGAGATTTGATATATATATCATACGATCTCACAGCAAGAGAAA	4228		
Db 589	ATTTTGGAGAAAGACACTTAATGACGATATATTTTAATAGTTTGGTAAAGATGTTCAA	648		

```

Oy 3869 ACTGTTGCTAAAGTATTGAGTAAGAAATAATCTCTCTATTATATTCATAGATGATCAT 3928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 AATGTTGATTTAGCTAGTGAAGAACTAAGAGAAAAATTTCTTGACTATTAGAGAAAGCA 592
Oy 3929 ACTGAACCTTGCTACAGACATCGGAAATGATAGCAATTCACAGATCCGACATTCATCG 3988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 AATACATTAAGTGAAGTGTGATGATTAATTAATAATTTAGAGAAAGCCGAAGATATTAAG 652
Oy 3969 AAAAATGGAATATATCACCAGTATTAAATGACAAAAATCTGAATATATCCAAAAACACATT 4048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 653 GAAATATATCTTATTAGTAATATGGAAGAACCAAAAGAAAAATATATTGACAAATTTATTA 712
Oy 4049 GAAACTATCTTGCTGATTAAGATGAGATGGATTAATTTGAAACGTTTATGTTGTAAT 4108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 713 AATATATTTGSCACAAAATTTGCAAAAAACMAAGAGTATACAGAAATGTACACAGTCAGT 772
Oy 4109 GAGAATGTGATTAAATGACGATGACATTTGCGAAGCTAATCCACTACAGATGAAATTAAT 4168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 773 GATGAACCTTTTAAATGATTTATTAATAATAGTTAGATGTTAATGAGAAATGAAAAAT 832
Oy 4169 GATGTTGATGATGATGAGAGCTTTTGATATATATCATGACATGTCACAGACAAAAGAGAA 4228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 833 ATTTTGGAGGAAGCTCAAGTTAAAGACGATATTTTATAGTTAGTAAAAAGTGTCAA 892
Oy 4229 TACACATTTGGAAGAAGAGTTAAACGAAAAAATCTGCTACTAAACATTCACATTGA 4284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 893 CAAAGAACACAAACCAATGTTGAAGCAAAAAAGTTGAAGAAAGTGTAGAGAAAGATGA 948

RESULT 6
US-09-410-464-9
: Sequence 9, Application US/09410464
: Patent No. 6395892
: GENERAL INFORMATION:
: APPLICANT: Strauss et al.
: TITLE OF INVENTION: Floral, homeotic genes for manipulation of flowering in
: TITLE OF INVENTION: poplar and other plant species.
: FILE REFERENCE: 53375
: CURRENT APPLICATION NUMBER: US/09/410,464
: CURRENT FILING DATE: 1999-10-01
: EARLIER APPLICATION NUMBER: 09/287,700

```

EARLIER FILING DATE: 1999-04-06
 EARLIER APPLICATION NUMBER: 60/080,851
 EARLIER FILING DATE: 1998-04-06
 NUMBER OF SEQ ID NOS: 24
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 9
 LENGTH: 11485
 TYPE: DNA
 ORGANISM: *Populus balsamifera* subsp. *trichocarpa*
 US-09-410-464-9

Query Match 0.98; Score 55.6; DB 4; Length 11485;
 Best Local Similarity 44.68; Pred. No. 0.0034;
 Matches 348; Conservative 0; Mismatches 424; Indels 8; Gaps 3;

4468 CACAAATTCAGCTATATGATCCGGATGATTAAGTCATGATGCTGAAGTAGAGAA 4527
 130 CTATTAAGAGGCAATCTGTGATCGCTCGGCACAAAGCATATGATTAAGATTTCTGC 169
 4528 ATTATGATCAAAAGATGTTTACGAAGATTCATCCACCGGCTGAAGCCTATATC 4587
 190 TTTCATTAAGACAGATCTGATCGATCTGCTCTACCTCCCGTAAGAGTGTGG 249
 4588 TATGGGTGGGTACATCTGAGAAATGATCTCTCAAGAGTGTGTTGGAAATCAG 4647
 250 TTGTCATGGGTGATTAAGATCAAGACTAATCT--GATGGGTCTATTGACATACAA 306
 4648 TTGTTGTTGTCATGGACAGACAAAGAAATTTGATATGACCTTTTATGTTAG 4707
 307 AGCTAGGCTGTTGCAAAAGATATCTCAACATTAATGATGACATATGAGAAAT 366
 4708 TTCACCTGTTATGATCTTGTGACTATATGATTAATGACATTAATAGTTGAT 4767
 367 TGCCCGGTTGCAAAAGACTACTATCTGATCTTATGTCGATCTGATCTGCTCA 426
 4768 AATGACATTTCAATTTAGACGTCGAGTGGCGTATCTAAATGCTCTATTACTATTC 4827
 427 GTGCAATTTTCTCACTGATTTGTTAAATGCTCTTCTGATGAGATCTTCAAGAGA 486
 4828 AATTCATTTTATGCTTCTCTCTAAATGATCTTGAAGAAACCATTTGTTGTT 4887
 487 AGTTATGTGACACCTCCCTCGATTTATGATGACG--TGATATGTTTGAACCTT 544
 4888 ATTGAAGCTTCTGCTATGTTTAAAGACGTCGGTTTGAATGATATCACTATCAA 4947
 545 AAGAAAGCTTAAATTAATGTTCTCAAAAGACCCCGTCTGTTTGAAGATTTCTC 604
 4948 AAGATATTTGAAGACATTTGTTTACTCAAGTTTACACATGATGTTTATTTACAT 5007
 605 TATGTGATCTGCTCTGCTGATTTTCTAGCAGTATGATTTCTCTCTTTT---AT 661
 5008 TGAATATGAGAGGAGCAGTATATATTTAGTTTATGTTTATGATATTTTATGCT 5067
 662 TAAGTCAGTATGAGGCTGATCATCTGCTTTTATGTTTATGATATGATATTTAT 721
 5068 TGAAGTTCACAAAAGATTTATGATTTTGTGATCAATGAGATCATTTTGAAGT 5127
 722 TGGGATGACATTTGATGTTTATTTAGTCTGAAGACAAAGTTGGCTAGAGATTTGAAT 781
 5128 TAAAGTGTGTTGTAATCAATTTATCTGTTATGATTTGTTAAACCGAATGTG 5187
 782 GAAAGATTTGGTATCTTCAATATTTCTGGGATTTAGAGTACATACCTACAGG 841
 5188 TTATATTTTATCAAGAAATTTCTCAAGAAATTTACTTAAGATTTCAACATAGATA 5247
 842 TTACCTTTCTCAGTGAATATGTTGACATATTTCTGAGCAGACTGACTTACTGA 901

RESULT 7
 US-08-771-602D-1
 Sequence 1, Application us/08771602D
 Patent No. 5976795
 GENERAL INFORMATION:

APPLICANT: Voytas, Daniel F.
 APPLICANT: Zou, Sigde
 TITLE OF INVENTION: Retrotransposon and Methods
 NUMBER OF SEQUENCES: 51
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
 STREET: 5370 Manhattan Circle, Suite 201
 CITY: Boulder
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/771,602D
 FILING DATE: 20-DEC-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/010,869
 FILING DATE: 31-JAN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Ferber, Donna M.
 REGISTRATION NUMBER: 33,878
 REFERENCE/DOCKET NUMBER: 8-96
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 499-8080
 TELEFAX: (303) 499-8089
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6660 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: not relevant
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: *Saccharomyces paradoxus*
 STRAIN: NRRL Y-17217
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1441..6321
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 1228..6602
 OTHER INFORMATION: /function="retrotransposon"
 OTHER INFORMATION: /product="Ty5-6p"
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 1228..1478
 OTHER INFORMATION: /function="5' LTR of Ty5-6p"
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: 2852..4827
 OTHER INFORMATION: /function="Integrase region coding"
 OTHER INFORMATION: region of Ty5-6p"
 US-08-771-602D-1

Query Match 0.88; Score 54.6; DB 2; Length 6660;
 Best Local Similarity 45.28; Pred. No. 0.0077;
 Matches 241; Conservative 0; Mismatches 289; Indels 3; Gaps 1;
 4661 GCGACAGACAAAGAAATTTGATTTATGACCTTTTATGTTAGTGTACCTGTATA 4720
 4981 GCCAAGGACATACACAAAGGCTGTATGATGATCAAGAACTTTGACACATCAT 5040

Qy	4841	GTCTTCCTCCTAATCATGACCTTGGACAATAAACCATCTGGTTATGTAAGAGTCT	4900
Db	5161	CAACCCAGCCGATTTATTAAATGAAAGTAATCCCGACTTAATGTAAGGAACATAATAGCGCGT	5220
Qy	4901	GCTCATGGGTTAAACACAGTCGGGTTTGGATGGTATCACACATCAAAAGAGTATTGGAA	4960
Db	5221	ATGATATGAGACTCAGCAAGGCCCATTACTATGSAACGAAACATATCAACAATCTCTTCAA	5280
Qy	4961	GACATGCTTTTACTCAAGTTTTCACAAATGATGTTTATTTACATTTGAATATGAAGAG	5020
Db	5281	AAGATGTGTTTCTCCGACATGAAAGCGAACAATGCGTAT---ACTTGTTCCACATCT	5337
Qy	5021	GGATCAGTAATATTTAGGTTTATATGTTATATGTTGATGATATCTTATGGTTGGAAGTTCAAA	5080
Db	5338	GATGGTCCCATCTACATTTGCCCTATACGTAGACAGCATTTCTGTGTGCTCCCTCCCTCG	5397
Qy	5081	AAAGTTATTGATTAATTTTGGATCAATGAGAGCATTTTGAAGTTAAAGTGTTCGT	5140
Db	5398	AAATATATATGACAGGTTAAAGCAGAAACATACAGAGTTTATCTCAATGAGAGATCTAGT	5457
Qy	5141	GAAATATCAATTAATCTGTGTTATGTAATTTCGTAAACCGAATCTCGTTATAT	5193
Db	5458	AAAGTTGACAAATTCCTCGGCTTTAACAATTATTCATTTTCAAATGAGAGACT	5510

```

RESULT 9
US-09-370-838-151
: Sequence 151: Application US/09370838
: Patent No 6444425
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Mohamath, Roadoh
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
: FILE REFERENCE: 210121.475G1
: CURRENT APPLICATION NUMBER: US/09/370,838
: CURRENT FILING DATE: 1999-08-09
: EARLIER APPLICATION NUMBER: US 09/285,323
: NUMBER OF SEQ ID NOS: 289
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 151
: LENGTH: 3275
: TYPE: DNA
: ORGANISM: Homo sapien
: US-09-370-838-151

```

Query Match	0.8%	Score 54.4	DB 4	Length 3275
Best Local Similarity	59.9%	Pred. No. 0.0069		
Matches	91	Conservative	0	Mismatches 61
			Indels	0
			Gaps	0
QY 1211	TTACCAAGCACACAGAGAAACCAAAAGGAAACACAGAGAACTCCTGAAACAATCTCAA			1270
Db 3121	TTACATGCAAGCAA			3180
QY 1271	AAGAAAGACCTGAAATCAAGGAAGGAATTAAGAAACTCCAAATATAGTTAAACGTTAAA			1330
Db 3181	AAA			3240
QY 1331	GGTGAAGAAAGAAAGCAAAAGCAAAACCTTC			1362
Db 3241	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTTC			3272

RESULT 10
US-09-797-906-1
; Sequence 1, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: ZIANGHE YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY

TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
FILE REFERENCE: C1001151CIP
CURRENT APPLICATION NUMBER: US/09/7797, 906
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1798
TYPE: DNA
ORGANISM: Human
US-09-797-906-1

Query Match
Best Local Similarity 58.0%; Score 53.2; DB 4; Length 1798;
Pred. No. 0.011; Mismatches 68; Indels 0; Gaps 0;
Matches 94; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1197 ATCTATTGTTAGTTTACCGACGAGAGAAACCAAGAGAAACGAGAGACTCAG 1256
DB 1629 ATGTCCTTGCGAGTGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1688
DB 1257 TGGAACATCTCAAAAGAGAAAGCTGAATCAAGAAAGAAATAGAAATCCCAAT 1316
DB 1689 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1748
QY 1317 CAGATAACGATAAGGTGAAAAAGAAAAAGAAAAAGAAAAA 1358
DB 1749 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1790

RESULT 11
US-08-446-855A-1
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
TITLE OF INVENTION: Phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 NO. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match
Best Local Similarity 46.2%; Score 53.2; DB 2; Length 8920;
Pred. No. 0.017; Mismatches 288; Indels 7; Gaps 2;
Matches 253; Conservative 0; Mismatches 288; Indels 7; Gaps 2;

QY 3760 AATGTTGAGCAATCTAATGTTAAACAGAGAGCTGCTAAGCACTGCAATTCAGACGA 3819
DB 5611 AGATATTGTGATGTAATTAATTAAGCAAAATTAATTAATTAATTAATTAATTAAT 5670
QY 3820 AGTTCTCT--CAAGAGGAGAGAAATCTTAATGAACAACCTGATATAGTTAGTCTGC 3876
DB 5671 AGATTGTAGAAAAGAGATATGATGAGAAAATTAATTAATTAATTAATTAATTAAT 5730
QY 3877 TAAAGTTATTGAGAAATGAAAAAATCTCTCTTAATTAATTAATTAATTAATTAATTA 3936
DB 5731 GAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5790
QY 3937 TGTACAGACTCGGGAATATGATGCAATTCACACAGAAATCCGATTCATGAAAAATGA 3996
DB 5791 AACATCACTTCACACTTAATGATCAATTAATTAATTAATTAATTAATTAATTAATTA 5850
QY 3997 AATATCACCAGATTAATGAGAAAAATCTGAAATTAATTAATTAATTAATTAATTAAT 4056
DB 5851 AATAGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5910
QY 4057 CCTTCTGATAGAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4116
DB 5911 AAACCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5970
QY 4117 GATTATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4176
DB 5971 TAAGATGAAGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 6030
QY 4177 GATGATGAGAGATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4232
DB 6031 TATGATTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6090
QY 4233 CATTTGAGAAAGATTAACGAAAAATTCGCTGCTACTTAACATTCCTGATGACATG 4292
DB 6091 TAATTTAGAAATGATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6150
QY 4293 ATCCAGA 4300
DB 6151 TTTAAACA 6158

RESULT 12
US-09-150-741-1
Sequence 1, Application US/09150741
Patent No. 6183996
GENERAL INFORMATION:
APPLICANT: Stewart et al.
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
TITLE OF INVENTION: Synthetase II
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/150,741
FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PL6380
EARLIER FILING DATE: 1992-12-16
EARLIER APPLICATION NUMBER: AU93/00617
EARLIER FILING DATE: 1993-12-02
EARLIER APPLICATION NUMBER: 08/446,855
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 8920
TYPE: DNA
ORGANISM: Plasmodium falciparum
US-09-150-741-1

Query Match
0.8%; Score 53.2; DB 4; Length 8920;

2.

DD 6151 YYYAACA 6158

Patent NO. 5766597

COMINT: UNLIED
ZTP. 10036

FILING DATE: 11;

US-08-257-073-4

maldives 94; conserva

DD 2145 AAAAG

US-09-134-001C-322

Matches 18

0
F
C
C
C

OY	3999	TAACCAGATGATTTATGAAAATAATCTGAATTAATATCCAAAACCATTTGAAGATATGC	4058
Db	14045	CTAAGAAAAAGCTAATCTACTGTTAACACAGCATATGCTCAACTAATAACAAATCTTTGCTA	14104
OY	4059	TTGCTGTAAGAGATTGGATGTAATTTGAAACGTATATGTTGATGAAATTGAGATGTGA	4118
Db	14105	ACTATGACGATGATTAATGTCAGCGCCAAATTACATTAAATGAGATTAAGACAAAAAAGATG	14164
OY	4119	TTATATGACATGATTCGTGAGGCTAAATCCACTCCAGATGAAATAATGATGTTCCAGA	4178
Db	14165	CTTATATATACCGTGTATTACATATGCTAAACAAAGATGATTAATCATCTGATGTCAATCAG	14224
OY	4179	TGAATGAGACTTTGTTAATTAATCATAGCATGTACGAGCAAGAGAAATACACATTGG	4238
Db	14225	CACAACTTGATTCACGCTGAAATTAATAAGGTACCAAAAGAGTCMAATACGACTTAAATAATG	14284
OY	4239	AGAAAGAGATTAACGAAAAAATTCGCTGACTTAACATTCCTGATTAACAAC	4290
Db	14285	ATCTAAATGTATATGACAAATTCGCTGCAAGCTTAAAGAAAGATGCTAATATACAC	14336

SULT 15

-09-443-01A-27
 Sequence 27. Application US/09443041A
 Patent No. 6465717
 GENERAL INFORMATION:
 APPLICANT: Farnado, Onolayo O.
 APPLICANT: Orozco, Buddy
 APPLICANT: Rafalski, Antoni
 APPLICANT: Shen, Jennie
 TITLE OF INVENTION: Sterol Metabolism Enzymes
 FILE REFERENCE:
 CURRENT APPLICATION NUMBER: US/09/443,041A
 CURRENT FILING DATE: 1999-11-18
 PRIOR APPLICATION NUMBER: 60/109,283
 PRIOR FILING DATE: 1998-11-20
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 27
 LENGTH: 1447
 TYPE: DNA
 ORGANISM: Glycine max
 US-09-443-041A-27

Query Match	0.8%	Score 51.4;	DB 4;	Length 1447;
Best Local Similarity	55.2%;	Pred. No. 0.025;		
Matches 100;	Conservative	0;	Mismatches 81;	Indels 0;
				Gaps 0;

QY	1198	TCCTATTGTTAGTTTACCACACACAGACAACCAAAAGCAACCAACGAGGAGACTACT	1257
QY	1254	TCCTAAGGCTGTGTTGGTTGTTAAAGACAAAAGATGGCAATGGACGAAATAGGTGAGA	1313
QY	1258	GGACAACATCTCAAAAGAGAAGAACTGAAATCTAAGAAAGACAAATTAAGAAACATCCAAATTC	1317
Db	1314	AGACAAAAA	1373
QY	1318	AGATTAACGATTAACGTCGAAAAAGAAAAAGAAAAAAGAAAAAAGCTTCACTGGAAATGAAAAAC	1377
Db	1374	AA	1433
QY	1378	A	1434
Db	1434	A	1434

Search completed: April 27, 2003, 16:51:08
Job time : 575 secs

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:54:05 ; Search time 356.12 Seconds

(Without alignments)
11330.158 Million cell updates/sec

Title: US-09-430-590e-3

Perfect score: 6426
1 tgtgtgtgtgtgtcactattt.....agaaatttatattccatca 6426

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424.4	100.0	6426	21	AAAS7920
2	392	6.1	392	21	AAAS7936
3	388.8	6.1	392	21	AAAS7923
4	388.8	6.1	392	21	AAAS7927
5	388.8	6.1	392	21	AAAS7928
6	388.8	6.1	392	21	AAAS7931
7	388.8	6.1	392	21	AAAS7932
8	387.2	6.0	392	21	AAAS7925
9	387.2	6.0	392	21	AAAS7929
10	387.2	6.0	392	21	AAAS7933
11	385.6	6.0	392	21	AAAS7930

12	384	6.0	392	21	AAAS7934	Candida albicans s
13	384	6.0	392	21	AAAS7935	Candida albicans s
14	382.4	6.0	392	21	AAAS7924	Candida albicans s
15	380.8	5.9	392	21	AAAS7926	Candida albicans s
16	328	5.1	1385	21	AAAS7989	Candida albicans s
17	280	4.4	1249	21	AAAS7947	Candida albicans s
18	280	4.4	183	21	AAAS7947	Candida albicans s
19	280	4.4	5611	21	AAAS7948	Candida albicans s
20	275.2	4.3	879	21	AAAS7991	Candida albicans s
21	264.4	4.1	454	21	AAAS7967	Candida albicans s
22	255.6	4.0	974	21	AAAS7992	Candida albicans s
23	252.8	3.9	1308	21	AAAS7949	Candida albicans s
24	249.8	3.9	469	21	AAAS7994	Candida albicans s
25	224.8	3.5	9850	21	AAAS8018	Candida albicans s
26	224.6	3.5	3604	21	AAAS7946	Candida albicans s
27	190	3.0	690	21	AAAS7951	Candida albicans s
28	174.4	2.7	936	22	AAAS8252	Candida albicans s
29	174.4	2.7	936	22	AAAS8254	Candida albicans s
30	174.4	2.7	936	22	AAAS8257	Candida albicans s
31	174.4	2.7	936	22	AAAS8259	Candida albicans s
32	174.4	2.7	936	22	AAAS8262	Candida albicans s
33	174.4	2.7	936	22	AAAS8255	Candida albicans s
34	172.2	2.7	936	22	AAAS8252	Candida albicans s
35	172.2	2.7	936	22	AAAS8254	Candida albicans s
36	172.2	2.7	936	22	AAAS8257	Candida albicans s
37	172.2	2.7	936	22	AAAS8259	Candida albicans s
38	172.2	2.7	936	22	AAAS8262	Candida albicans s
39	172.2	2.7	936	22	AAAS8255	Candida albicans s
40	170.4	2.7	1912	21	AAAS7952	Candida albicans s
41	164	2.6	1672	21	AAAS7950	Candida albicans s
42	130	2.0	138	21	AAAS7937	Candida albicans s
43	104.2	1.6	770	21	AAAS7964	Candida albicans s
44	104.2	1.6	951	21	AAAS8017	Candida albicans s
45	100.6	1.6	1348	21	AAAS7983	Candida albicans s

ALIGNMENTS

RESULT 1
ID AAAS7920 standard; DNA: 6426 BP.
XX
AC AAAS7920;
XX
DT 11-OCT-2000 (first entry)
XX
DE Candida albicans pcal retrotransposon.
KW Retrotransposon; pcal; Tca2; Tyl; copia; long terminal repeat; LTR;
KW gag gene; group antigen; polypeptide; pol; aspartate protease; integrase;
KW reverse transcriptase; Rnsen; pseudoknot; readthrough translation;
KW stop codon suppression; gene delivery; gene therapy vector;
KW genetic vaccine composition; immunogenic; transgenic animal; ds.
XX
OS Candida albicans strain hog1042.
XX
FH Key
FH Location/Qualifiers
FT 1..280
FT /tag= a
FT repeat_unit
FT 1..6
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Imperfect 6 bp repeat"
FT 120..125
FT /tag= c
FT TATA_signal
FT /standard name= "TATA box"
FT 185..190
FT /tag= d
FT /standard name= "TATA box"
FT 201..206
FT /tag= e
FT repeat_unit
FT 275..280

```

FT      /tag= f
FT      /rpt_type= INVERTED
FT      /note= "Imperfect 6 bp repeat"
FT      281..291
FT      /tag= g
FT      /note= "this site in the corresponding RNA is a primer
FT      binding site for reverse transcriptase-mediated
FT      minus-strand DNA synthesis"
FT
FT      CDS
FT      498..1372
FT      /tag= h
FT      /product= "Gag (group antigen) protein"
FT      /codon= (seq:"ctg", aa:Ser)
FT      /note= "5 serine residues in the gag protein are encoded
FT      by this non-standard Ser codon"
FT      498..6103
FT      /tag= i
FT      /product= "Gag (group antigen)-pol (polyprotein)
FT      readthrough protein"
FT      /note= "The gag-pol readthrough protein is the result of
FT      stop codon suppression mediated by pseudoknot
FT      formation in the mRNA"
FT      /codon= (seq:"ctg", aa:Ser)
FT      /note= "12 serine residues in the gag-pol readthrough
FT      protein are encoded by this non-standard Ser
FT      codon"
FT      /transl_except= (pos:1367..1372, aa:Glu)
FT      1381..1455
FT      /tag= j
FT      /note= "The RNA corresponding to this region forms a
FT      pseudoknot, allowing gag ORF stop codon
FT      suppression and translation of the gag-pol
FT      readthrough protein"
FT      1373..6100
FT      /tag= k
FT      /product= "Pol (polyprotein), comprising aspartate
FT      protease, integrase, reverse transcriptase
FT      (RT) and RNaseH"
FT      /codon= (seq:"ctg", aa:Ser)
FT      /note= "7 serine residues in the pol protein are encoded
FT      by this non-standard Ser codon"
FT      3455..3465
FT      /tag= l
FT      /note= "Polypurine tract 2 (PPT2)"
FT      /complement= (3455..3465)
FT      /tag= m
FT      /note= "this site in the corresponding RNA is a primer
FT      binding site for reverse transcriptase-mediated
FT      plus-strand DNA synthesis"
FT      6136..6146
FT      /tag= n
FT      /note= "Polypurine tract 1 (PPT1)"
FT      /complement= (6136..6146)
FT      /tag= o
FT      /note= "this site in the corresponding RNA is a primer
FT      binding site for reverse transcriptase-mediated
FT      plus-strand DNA synthesis"
FT
FT      primer_bind
FT      6147..6426
FT      /tag= p
FT      repeat_unit
FT      6147..6152
FT      /tag= q
FT      /rpt_type= INVERTED
FT      /note= "Imperfect 6 bp repeat"
FT      6266..6271
FT      /tag= r
FT      /standard name= "TATA box"
FT      6331..6336
FT      /tag= s
FT      /standard name= "TATA box"
FT      6347..6351
FT      /tag= t
FT      polyA_signal
FT      6421..6426
FT      /tag= u
FT      repeat_unit
FT      /rpt_type= INVERTED

```

```

FT      /note= "Imperfect 6 bp repeat"
FT      XX
FT      FN      WO200026397-A1.
FT      XX
FT      PD      11-MAY-2000.
FT      XX
FT      PF      01-NOV-1999; 99WO-N200179.
FT      PR      30-OCT-1998; 98CA-2249046.
FT      PR      30-OCT-1998; 98US-0106342.
FT      XX
FT      PA      (JANCO) JANSEN PHARM NV.
FT      PI      Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;
FT      DR      WPI: 2000-365640/31.
FT      DR      P-FSDB: AAB03126, AAB03127, AAB03128.
FT      PT
FT      PT      Novel retrotransposon expression vectors useful for expressing an
FT      PT      antigen, epitope or therapeutic agent, or detecting genes or the
FT      PT      presence of Candida in a sample -
FT      PS
FT      PS      Claim 12; Fig 2B; 204pp; English.
FT      XX
FT      CC      The invention relates to novel retrotransposons from the yeast Candida
FT      CC      albicans which have a copy number of 40-150, preferably 50-100 copies
FT      CC      per genome. In particular, the invention relates to the novel C.
FT      CC      albicans Ty1/copia retrotransposon pcal (AAB57920), and to the integrated
FT      CC      form of this retrotransposon, designated rca2, and to the novel C.
FT      CC      albicans retrotransposons 1-28. pcal was initially isolated from C.
FT      CC      albicans HOG1042 and has a copy number of 50-100 copies per cell. It
FT      CC      comprises identical 280 bp long terminal repeats (LTRs) and two open
FT      CC      reading frames (ORFs). The first ORF encodes a gag (gag) consisting of an
FT      CC      protein, and the second ORF encodes a polyprotein (pol) consisting of an
FT      CC      aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
FT      CC      gag and pol ORFs of pcal are in the same reading frame, separated only by
FT      CC      a termination codon (TGA). Translation of the pol ORF occurs through
FT      CC      the occasional readthrough suppression of the stop codon, which is
FT      CC      mediated by the formation of a pseudoknot within the gag-pol mRNA.
FT      CC      The retrotransposons of the invention can be used as vectors for in
FT      CC      vitro or in vivo transformation and expression. They can thus be used
FT      CC      for the delivery and expression of a therapeutic, immunological or
FT      CC      immunogenic molecule (e.g., an antigen) and may also be used for
FT      CC      eliciting an immunological response in a host organism. They are
FT      CC      therefore useful in genetic vaccine compositions and for gene therapy,
FT      CC      particularly where the use of retroviral vectors is unsafe or
FT      CC      undesirable. Additionally, the retrotransposons may be used to generate
FT      CC      transgenic animals, to detect the presence of Candida in a sample, to
FT      CC      detect and disrupt genes, and to assign functions to nucleotide
FT      CC      sequences. The present sequence represents the C. albicans pcal
FT      CC      retrotransposon.
FT      XX
FT      XX      Sequence 6426 BP; 2228 A; 1032 C; 1258 G; 1908 T; 0 other:
FT      SQ
FT
FT      Query Match      100.0%; Score 6424.4; DB 21; Length 6426;
FT      Best Local Similarity 100.0%; Pred. No. 0;
FT      Matches 6425; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
FT
FT      QY      1 tgttggttggtgactatttgggtgcaagaactgataatgaagaatgattataga 60
FT      DB      1 tgttggttggtgactatttgggtgcaagaactgataatgaagaatgattataga 60
FT      QY      61 gaatggaatttttcacacacacgagtgatgacgaactaactatgtgtagt 120
FT      DB      61 gaatggaatttttcacacacacgagtgatgacgaactaactatgtgtagt 120
FT      QY      121 ataataaggtatgaataccacacatccagagataatcaagagatagaaggagagt 180
FT      DB      121 ataataaggtatgaataccacacatccagagataatcaagagatagaaggagagt 180
FT      QY      181 tcaatatatactctgtgaataacttggttctaattcaactatacaactagaagct 240

```

Db 181 tcaatataatctgtgtgaataaactcgttctaattcactatatacaacaaagtagcgtgt 240
QY 241 acaagctcaatctcaggtaaaagaagtttatctcactcgaatagagaagtcgaatagat 300
Db 241 acaagctcaatctcaggtaaaagaagtttatctcactcgaatagagaagtcgaatagat 300
QY 301 aatcaattcgtcccaaatatagcgttgtataaatcaatcagtcctcagatttgtatattgatt 360
Db 301 aatcaattcgtcccaaatatagcgttgtataaatcaatcagtcctcagatttgtatattgatt 360
QY 361 gatagttcgaagtttgaaagttacagaatttcacaagaatagattccgcaagaatagtga 420
Db 361 gatagttcgaagtttgaaagttacagaatttcacaagaatagattccgcaagaatagtga 420
QY 421 taacgaagggaaggttcatggaaggtgtgtgaatgaagcgaatgctatgtaaggtgga 480
Db 421 taacgaagggaaggttcatggaaggtgtgtgaatgaagcgaatgctatgtaaggtgga 480
QY 481 acatataaagctgaatcacaatatacgtcttcataaaatctaattgaacttacctaagtgc 540
Db 481 acatataaagctgaatcacaatatacgtcttcataaaatctaattgaacttacctaagtgc 540
QY 541 cgtcgtgaatcaagaagaagcgtgtgaataatgaagaaatctaatttcacagttgc 600
Db 541 cgtcgtgaatcaagaagaagcgtgtgaataatgaagaaatctaatttcacagttgc 600
QY 601 ttaccccgatgttttgaaattttgtcctgactaatacctcaagaagataatcacaagttaa 660
Db 601 ttaccccgatgttttgaaattttgtcctgactaatacctcaagaagataatcacaagttaa 660
QY 661 aaaggtgaaggaatattatcttctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
Db 661 aaaggtgaaggaatattatcttctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
QY 721 tgaatagttcaagttgcatgcatgtttctaagcaccagaagcacttgcaaaaggaagcaaa 780
Db 721 tgaatagttcaagttgcatgcatgtttctaagcaccagaagcacttgcaaaaggaagcaaa 780
QY 781 cttaatacaagcgtcgtctatgactgtctgttactaaactcaagaatatacctaactagtaa 840
Db 781 cttaatacaagcgtcgtctatgactgtctgttactaaactcaagaatatacctaactagtaa 840
QY 841 gatcttgctgaagtttgtaagcttgtaacatgagttagtggttggttataacttcata 900
Db 841 gatcttgctgaagtttgtaagcttgtaacatgagttagtggttggttataacttcata 900
QY 901 ttgctgcaaggttggaagaagaacttgagaagaatacctcaacaacttcaaacgtgtgtga 960
Db 901 ttgctgcaaggttggaagaagaacttgagaagaatacctcaacaacttcaaacgtgtgtga 960
QY 961 tgaagtatgctgaagttcctccaatctcactaaggttcaagcttgtaacttaatcaatgtga 1020
Db 961 tgaagtatgctgaagttcctccaatctcactaaggttcaagcttgtaacttaatcaatgtga 1020
QY 1021 gaaatcagaagccttaagttgttttgtaatacttgctcactactactactactactactact 1080
Db 1021 gaaatcagaagccttaagttgttttgtaatacttgctcactactactactactactactact 1080
QY 1081 tcaagctgacaatgatacatcagtaactccaagtgtgtcctacatagctggaagaatgtgt 1140
Db 1081 tcaagctgacaatgatacatcagtaactccaagtgtgtcctacatagctggaagaatgtgt 1140
QY 1141 tgaatcctcgtatgatagtatgtgtgtgacattcacaagaacaaataatgaacttaact 1200
Db 1141 tgaatcctcgtatgatagtatgtgtgtgacattcacaagaacaaataatgaacttaact 1200
QY 1201 tatgtgtatgttaccagcaccagagaacaaagaagaaacccagagagaaactcacttga 1260
Db 1201 tatgtgtatgttaccagcaccagagaacaaagaagaaacccagagagaaactcacttga 1260
QY 1261 acaatctcaaaagaagaacttgataatcaagaagaagaatgaagaacatccaaatcaga 1320
Db 1261 acaatctcaaaagaagaacttgataatcaagaagaagaatgaagaacatccaaatcaga 1320

QY 1321 taacgaataagttgaaaaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1380
Db 1321 taacgaataagttgaaaaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1380
QY 1381 tgcgtctctatgaattgtgtatgataatcactaatctcagaacaaacagcttccagtt 1440
Db 1381 tgcgtctctatgaattgtgtatgataatcactaatctcagaacaaacagcttccagtt 1440
QY 1441 agaaaatctcactctccttaagtcttcttgaaacgttaatgaattttaaaggttgaagtt 1500
Db 1441 agaaaatctcactctccttaagtcttcttgaaacgttaatgaattttaaaggttgaagtt 1500
QY 1501 taacaagatctcagtgatgatacagtcggtgcacaaatctcgttgtgaacaaataagatat 1560
Db 1501 taacaagatctcagtgatgatacagtcggtgcacaaatctcgttgtgaacaaataagatat 1560
QY 1561 atgtcgtgaatgttgaaggaagcacaatgtgaattctgtgtgtgtgtgtgtgtgtgtgtgt 1620
Db 1561 atgtcgtgaatgttgaaggaagcacaatgtgaattctgtgtgtgtgtgtgtgtgtgtgtgt 1620
QY 1621 agcagattgtatgtgtatctataatcaagagtcggtatgtgtcgaattcagtagaaga 1680
Db 1621 agcagattgtatgtgtatctataatcaagagtcggtatgtgtcgaattcagtagaaga 1680
QY 1681 taacattgataccagaagaagtccttaactctgttgaggtttgaaacaaattggaagagc 1740
Db 1681 taacattgataccagaagaagtccttaactctgttgaggtttgaaacaaattggaagagc 1740
QY 1741 aggaattgaatgtcttatactataaagaalcaagtgtatgtatcttaacccaataatgtgtcc 1800
Db 1741 aggaattgaatgtcttatactataaagaalcaagtgtatgtatcttaacccaataatgtgtcc 1800
QY 1801 tactatcttgcttgaagaagaagaagtcgtcgtatcttataatgtgtgtgtgtgtgtgtgtgt 1860
Db 1801 tactatcttgcttgaagaagaagaagtcgtcgtatcttataatgtgtgtgtgtgtgtgtgtgt 1860
QY 1861 agaaatcttgaagaatgtgtatcttgaatgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1920
Db 1861 agaaatcttgaagaatgtgtatcttgaatgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1920
QY 1921 ccaagaatgaagaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1980
Db 1921 ccaagaatgaagaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1980
QY 1981 tagtctcgaagcattcaaatcttcttgacgaaggttgatgtttataggtgtgaatttc 2040
Db 1981 tagtctcgaagcattcaaatcttcttgacgaaggttgatgtttataggtgtgaatttc 2040
QY 2041 cccataatgagttgaacaattgtctacaaactgagagaatagaagaatattatattccaa 2100
Db 2041 cccataatgagttgaacaattgtctacaaactgagagaatagaagaatattatattccaa 2100
QY 2101 ttgtatgtcaaatctatgtcatttgagaagaatctgttgttatacaaaaataccaaggtct 2160
Db 2101 ttgtatgtcaaatctatgtcatttgagaagaatctgttgttatacaaaaataccaaggtct 2160
QY 2161 cgtactcaacacttcaaaaagagctctcaaaaagattcgtatgtaaggtatgtcatt 2220
Db 2161 cgtactcaacacttcaaaaagagctctcaaaaagattcgtatgtaaggtatgtcatt 2220
QY 2221 atcgaatgtccaaacagagaagtcacaatcattcagaagaagaagaagcctcgagaagaca 2280
Db 2221 atcgaatgtccaaacagagaagtcacaatcattcagaagaagaagaagcctcgagaagaca 2280
QY 2281 tgaagagactcaattgtgatactcgtgcattgaagtcgaagaataaagaatgtgtatt 2340
Db 2281 tgaagagactcaattgtgatactcgtgcattgaagtcgaagaataaagaatgtgtatt 2340
QY 2341 aagctcgttataagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2400
Db 2341 aagctcgttataagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2400

QY 2401 ggttaagagctcttaattcaacgataaagcttggaataatcggtttaacgataagt 2460
 Db 2401 ggttaagagctcttaattcaacgataaagcttggaataatcggtttaacgataagt 2460
 QY 2461 ggcatactcaagaagtataatgctcttgatcccaacaccttgatataatgct 2520
 Db 2461 ggcatactcaagaagtataatgctcttgatcccaacaccttgatataatgct 2520
 QY 2521 cggatattggaggagactaagcgcatactgctcgtgaacttaagtctcgccgagt 2580
 Db 2521 cggatattggaggagactaagcgcatactgctcgtgaacttaagtctcgccgagt 2580
 QY 2581 tgttaataaatgttattcaacagattcaagagatcggttgacaccttggtcccaaat 2640
 Db 2581 tgttaataaatgttattcaacagattcaagagatcggttgacaccttggtcccaaat 2640
 QY 2641 actcaattgattattatgtatgtatcaattctcatcaatgatcaacacactccag 2700
 Db 2641 actcaattgattattatgtatgtatcaattctcatcaatgatcaacacactccag 2700
 QY 2701 tcgttcaactcaaggagacaaacccctatggttgctattatcaatlaagtgaggaaatt 2760
 Db 2701 tcgttcaactcaaggagacaaacccctatggttgctattatcaatlaagtgaggaaatt 2760
 QY 2761 ctaccggtttccctttgcacatgctgtgcgttaattagtaattgcacacgaaaaagaa 2820
 Db 2761 ctaccggtttccctttgcacatgctgtgcgttaattagtaattgcacacgaaaaagaa 2820
 QY 2821 ccgttaccgaagtacatcaacataaaggagctccctcatcgatcaatgggtgctgattg 2880
 Db 2821 ccgttaccgaagtacatcaacataaaggagctccctcatcgatcaatgggtgctgattg 2880
 QY 2881 ctacgctagagattggttttagttattacggttgcttaaaaaatacggtgctgattat 2940
 Db 2881 ctacgctagagattggttttagttattacggttgcttaaaaaatacggtgctgattat 2940
 QY 2941 ccttagccctaatgtccgtatatattcggaagctatgaagttatcaacccctatccaaaaa 3000
 Db 2941 ccttagccctaatgtccgtatatattcggaagctatgaagttatcaacccctatccaaaaa 3000
 QY 3001 ctatccactacactatgltacacatgltcctatggtctgaagttatccagggaaagca 3060
 Db 3001 ctatccactacactatgltacacatgltcctatggtctgaagttatccagggaaagca 3060
 QY 3061 actgggagctcaatgaaggtacgcggaaacatgtgtgaaagtgaatatgaataacaa 3120
 Db 3061 actgggagctcaatgaaggtacgcggaaacatgtgtgaaagtgaatatgaataacaa 3120
 QY 3121 tgaagtgtgacatgcccagaaggtatctcaagttcaagccacatcggtttacttaac 3180
 Db 3121 tgaagtgtgacatgcccagaaggtatctcaagttcaagccacatcggtttacttaac 3180
 QY 3181 tacgggtacagtttcaacgaatalgtatataatgatactccagtlacagattaccattga 3240
 Db 3181 tacgggtacagtttcaacgaatalgtatataatgatactccagtlacagattaccattga 3240
 QY 3241 gaatccgaatgatttttctaaaccttcaactaactgaagaatccacagatatgctatc 3300
 Db 3241 gaatccgaatgatttttctaaaccttcaactaactgaagaatccacagatatgctatc 3300
 QY 3301 cgaagtaaatcgatgtgaatgaaatccaaacccaggtcccaagcttaaaacccctgggataa 3360
 Db 3301 cgaagtaaatcgatgtgaatgaaatccaaacccaggtcccaagcttaaaacccctgggataa 3360
 QY 3361 tccggtgtctaaacctctcaactgtgtaccagagacttcaataaggaagcttaaga 3420
 Db 3361 tccggtgtctaaacctctcaactgtgtaccagagacttcaataaggaagcttaaga 3420
 QY 3421 gacctattcaaaacacacaaagagagcccttccatccaggggagggacaaacgctt 3480
 Db 3421 gacctattcaaaacacacaaagagagcccttccatccaggggagggacaaacgctt 3480
 QY 3481 ggaatctactgctcaggttgagatactacacaaacccagactggtactccggttcgga 3540

Db 3481 ggaatctactgctcaggttgagatactacacaaacccagactggtactccggttcgga 3540
 QY 3541 ggaagcaaatgttcaaggaaacagatcattcggttgacgtgtttaaagaacagcttc 3600
 Db 3541 ggaagcaaatgttcaaggaaacagatcattcggttgacgtgtttaaagaacagcttc 3600
 QY 3601 aagaatgttgatcttctgatactccgaacatgctgtcgaagtgaatgaacgcaaatcc 3660
 Db 3601 aagaatgttgatcttctgatactccgaacatgctgtcgaagtgaatgaacgcaaatcc 3660
 QY 3661 cctgttactggtctaaagaatcggttgaactgaagaataatagatgaggagaataatttc 3720
 Db 3661 cctgttactggtctaaagaatcggttgaactgaagaataatagatgaggagaataatttc 3720
 QY 3721 atttccgggggttgatgatactgctgcgtgatacactcaaatglttgagcaactaatgt 3780
 Db 3721 atttccgggggttgatgatactgctgcgtgatacactcaaatglttgagcaactaatgt 3780
 QY 3781 tgaacacagggatgctgtgtgaacagtcgaatccaaagcgaagtttctcaaggaggaaagat 3840
 Db 3781 tgaacacagggatgctgtgtgaacagtcgaatccaaagcgaagtttctcaaggaggaaagat 3840
 QY 3841 acttaatgacaaactgataatgattgatactggttctaaagttatgtgaatgaaataat 3900
 Db 3841 acttaatgacaaactgataatgattgatactggttctaaagttatgtgaatgaaataat 3900
 QY 3901 cttccctaatattcaatgataatgatacactgaactgcttaagagctcgggaatgataag 3960
 Db 3901 cttccctaatattcaatgataatgatacactgaactgcttaagagctcgggaatgataag 3960
 QY 3961 caattcaacagatccgaatcattcgaacaaatgaataatcacaggtgataatgaa 4020
 Db 3961 caattcaacagatccgaatcattcgaacaaatgaataatcacaggtgataatgaa 4020
 QY 4021 aaatactgaataatlocaaaacacatltgaagttatcctgtgtgaataagagattgata 4080
 Db 4021 aaatactgaataatlocaaaacacatltgaagttatcctgtgtgaataagagattgata 4080
 QY 4081 atttgaacagttatattgttgataatgataatgtgaattgaatgacgttgacatgcga 4140
 Db 4081 atttgaacagttatattgttgataatgataatgtgaattgaatgacgttgacatgcga 4140
 QY 4141 agctaatccactacacagatgaataatgataatgttcaagttgaatgataatgataa 4200
 Db 4141 agctaatccactacacagatgaataatgataatgttcaagttgaatgataatgataa 4200
 QY 4201 tcatagcatgtcagcagcaagaagaatalacacatttgaagaagaagttaacgaaat 4260
 Db 4201 tcatagcatgtcagcagcaagaagaatalacacatttgaagaagaagttaacgaaat 4260
 QY 4261 tgcgtgtactaaacttcaactgtgatacactgatacctcaagaagaagaatcagaagtgataa 4320
 Db 4261 tgcgtgtactaaacttcaactgtgatacactgatacctcaagaagaagaatcagaagtgataa 4320
 QY 4321 tactgttgaacccaagaagatcgaaacccaagaagaaggtggtccatcaactgataat 4380
 Db 4321 tactgttgaacccaagaagatcgaaacccaagaagaaggtggtccatcaactgataat 4380
 QY 4381 aaacaaagaatcgaaataacagttacacatagtttaaagaagtgtgtagaagcgttatataa 4440
 Db 4381 aaacaaagaatcgaaataacagttacacatagtttaaagaagtgtgtagaagcgttatataa 4440
 QY 4441 ccccaagaggtatttatacagcgttcgtaacaaatcgactaataatgatacgggattgat 4500
 Db 4441 ccccaagaggtatttatacagcgttcgtaacaaatcgactaataatgatacgggattgat 4500
 QY 4501 aaagtcaatgaatgagtaagtaagaatattagatcaaaagattttacgaagaagttcc 4560
 Db 4501 aaagtcaatgaatgagtaagtaagaatattagatcaaaagattttacgaagaagttcc 4560
 QY 4561 aattccacaggtgtgaagctatatactatggttgggtacatactgagaataattgatac 4620

Db 4561 aattccaccggtgtgaagccatatactatgtgtgtgtatcacatgagaattgtctc 4620
 QY 4621 tctcaagggtgtgttcggaaatcacgtgtgtgtccatgagcaacgacaaaaggaaa 4680
 Db 4621 tctcaagggtgtgttcggaaatcacgtgtgtccatgagcaacgacaaaaggaaa 4680
 QY 4681 atgtgatataacccttttagtcttagttcacgcttatagatcttctgtactaaatt 4740
 Db 4681 atgtgatataacccttttagtcttagttcacgcttatagatcttctgtactaaatt 4740
 QY 4741 atgtgaataataggtgtgtgaattaggaaatgaacaattagaagctgcagtcgc 4800
 Db 4741 atgtgaataataggtgtgtgaattaggaaatgaacaattagaagctgcagtcgc 4800
 QY 4801 gtatcctaattgaccttatctatctatccaattccaattatgtcttctccctaatt 4860
 Db 4801 gtatcctaattgaccttatctatctatccaattccaattatgtcttctccctaatt 4860
 QY 4861 accttgaagaaacacatctgtgtgtatgaacgcttcgtctatgtgttaaacagtc 4920
 Db 4861 accttgaagaaacacatctgtgtgtatgaacgcttcgtctatgtgttaaacagtc 4920
 QY 4921 gggtgttgatgtgtatcacatccaagaagatgtgaagacatgtgtttctcaagt 4980
 Db 4921 gggtgttgatgtgtatcacatccaagaagatgtgaagacatgtgtttctcaagt 4980
 QY 4981 tttaacaatgattgtttatcttccatgtgatatgaagaggtgcgttatattagg 5040
 Db 4981 tttaacaatgattgtttatcttccatgtgatatgaagaggtgcgttatattagg 5040
 QY 5041 tttaatagttgatgtatcttctatgtgttggaagttcacaaaagtatltgataatttgc 5100
 Db 5041 tttaatagttgatgtatcttctatgtgttggaagttcacaaaagtatltgataatttgc 5100
 QY 5101 ggtatcaattgagagatcatcttgaagttaagtgtgtgtggaatatccaattatctctg 5160
 Db 5101 ggtatcaattgagagatcatcttgaagttaagtgtgtgtggaatatccaattatctctg 5160
 QY 5161 tattgaatttcgttaaaacgcaatctgttatattatctccaagaaatttctcaagaa 5220
 Db 5161 tattgaatttcgttaaaacgcaatctgttatattatctccaagaaatttctcaagaa 5220
 QY 5221 attactaaagatttcaaaatagatgactcatactggaaaaaacataccctgtctcgaa 5280
 Db 5221 attactaaagatttcaaaatagatgactcatactggaaaaaacataccctgtctcgaa 5280
 QY 5281 tgaacaatatgaaaaggtgtccaataatcgtgaaaacgtttaaaccagaaatgttttga 5340
 Db 5281 tgaacaatatgaaaaggtgtccaataatcgtgaaaacgtttaaaccagaaatgttttga 5340
 QY 5341 aaaggttcggaatgagacatctgtctgaacctgtatgctaaaaaacttaccaaagtggtc 5400
 Db 5341 aaaggttcggaatgagacatctgtctgaacctgtatgctaaaaaacttaccaaagtggtc 5400
 QY 5401 tggcgtgtcttatactggtgtgcacaaacacagtlccagatalatcgtctgtagtaattc 5460
 Db 5401 tggcgtgtcttatactggtgtgcacaaacacagtlccagatalatcgtctgtagtaattc 5460
 QY 5461 gttggtgttcaaatctgcaaatccaattgccaattgtatgtgaatttattatgtctc 5520
 Db 5461 gttggtgttcaaatctgcaaatccaattgccaattgtatgtgaatttattatgtctc 5520
 QY 5521 taggtatatacaaaatagatgagatalacattgagtaacaaagaagaagttgaatac 5580
 Db 5521 taggtatatacaaaatagatgagatalacattgagtaacaaagaagaagttgaatac 5580
 QY 5581 accaccaaaatattgttatggaatgttcagtgatgcgttcaattgtgacagagattga 5640
 Db 5581 accaccaaaatattgttatggaatgttcagtgatgcgttcaattgtgacagagattga 5640
 QY 5641 taagaataatctattagtggaactgtgattgtgaattgaaatttgggtcgaatggcgac 5700
 Db 5641 taagaataatctattagtggaactgtgattgtgaattgaaatttgggtcgaatggcgac 5700

QY 5701 caaaaaacaaacggtatagacacaaagctcagaagctgtgtgaatgtgtgtcttaatta 5760
 Db 5701 caaaaaacaaacggtatagacacaaagctcagaagctgtgtgaatgtgtgtcttaatta 5760
 QY 5761 tacaatgttgaagctatagataataaaaaacattatgatttgggtttggaatgg 5820
 Db 5761 tacaatgttgaagctatagataataaaaaacattatgatttgggtttggaatgg 5820
 QY 5821 taagatactgttcatcaagacacaaacgctgtgtatgaatttgaagaataactatgt 5880
 Db 5821 taagatactgttcatcaagacacaaacgctgtgtatgaatttgaagaataactatgt 5880
 QY 5881 tcaaccaatcgaacaaatagatatctgtctataagtttctacgccaattgataatgaa 5940
 Db 5881 tcaaccaatcgaacaaatagatatctgtctataagtttctacgccaattgataatgaa 5940
 QY 5941 agtatttcaatactatctatgttgaagacaaatgataattacgacctgttatgactaagt 6000
 Db 5941 agtatttcaatactatctatgttgaagacaaatgataattacgacctgttatgactaagt 6000
 QY 6001 tctaaagctgtctaaatccaagcatcgtltgaaggtatgaataaagcttctgaactaga 6060
 Db 6001 tctaaagctgtctaaatccaagcatcgtltgaaggtatgaataaagcttctgaactaga 6060
 QY 6061 agataatcaaacacgtgatacaaaatgcaataagcgacgaataagtgtatctatcaact 6120
 Db 6061 agataatcaaacacgtgatacaaaatgcaataagcgacgaataagtgtatctatcaact 6120
 QY 6121 attatcgtatgtctcaatacgaagggaggtgtgtgtgtgacatttctgtcagaagactg 6180
 Db 6121 attatcgtatgtctcaatacgaagggaggtgtgtgtgtgacatttctgtcagaagactg 6180
 QY 6181 atcaatgaaatgattgtgtatbtatgagaatggaaaatttccatcacacatcagtgat 6240
 Db 6181 atcaatgaaatgattgtgtatbtatgagaatggaaaatttccatcacacatcagtgat 6240
 QY 6241 gacgaacataaactatattgtgtgtatgaataaaggtatgaataaacataccacagaa 6300
 Db 6241 gacgaacataaactatattgtgtgtatgaataaaggtatgaataaacataccacagaa 6300
 QY 6301 tatcaacgagatagaagggaggttccaatatatacttctgtgaataaactcgttct 6360
 Db 6301 tatcaacgagatagaagggaggttccaatatatacttctgtgaataaactcgttct 6360
 QY 6361 aattactatatacaacactagaacgtgtacacgctcaatctcaggttaagaagaattatatt 6420
 Db 6361 aattactatatacaacactagaacgtgtacacgctcaatctcaggttaagaagaattatatt 6420
 QY 6421 ccatca 6426
 Db 6421 ccatca 6426

RESULT 2
 AA57936
 ID AA57936 standard; DNA; 392 BP.
 AC AA57936;
 XX
 DT 10-OCY-2000 (first entry)
 XX
 DE Candida albicans strain p36 TCa2 retrotransposon 5' region.
 XX
 KW Retrotransposon; pCa1; TCa2; Ty1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 OS Candida albicans strain p36.
 XX
 PN WO200026397-A1.

XX 11-MAY-2000.
 PD
 XX
 XX 01-NOV-1999; 99WO-NZ00179.
 PF
 XX 30-OCT-1998; 98CA-2249046.
 PR 30-OCT-1998; 98US-0106342.
 XX
 XX (JANC) JANSSEN PHARM NV.
 PA
 PI Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;
 DR WPI; 2000-365640/31.
 XX
 XX Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -
 XX
 PS Example 9; Fig 9; 204pp; English.

The invention relates to novel retrotransposons from the yeast *Candida albicans* which have a copy number of 40-150, preferably 50-100 copies per genome. In particular, the invention relates to the novel *C. albicans* TY1/copia retrotransposon pCal (AA57920), and to the integrated form of this retrotransposon, designated TCa2, and to the novel *C. albicans* retrotransposons 1-28. pCal was initially isolated from *C. albicans* HOG1042 and has a copy number of 50-100 copies per cell. It comprises identical 280 bp long terminal repeats (LTRs) and two open reading frames (ORFs). The first ORF encodes a gag (group antigen) protein, and the second ORF encodes a polyprotein (pol) consisting of an aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and pol ORFs of pCal are in the same reading frame, separated only by a termination codon (TGA). Translation of the pol ORF occurs through the occasional readthrough suppression of the stop codon, which is mediated by the formation of a pseudoknot within the gag-pol mRNA. The retrotransposons of the invention can be used as vectors for in vitro or in vivo transformation and expression. They can thus be used for the delivery and expression of a therapeutic, immunological or immunogenic molecule (e.g., an antigen) and may also be used for eliciting an immunological response in a host organism. They are therefore useful in genetic vaccine compositions and for gene therapy, particularly where the use of retroviral vectors is unsafe or undesirable. Additionally, the retrotransposons may be used to generate transgenic animals, to detect the presence of *Candida* in a sample, to detect and disrupt genes, and to assign functions to nucleotide sequences. Sequences AA57923-AS7936 represents TCa2 retrotransposon 5' regions from a variety of *C. albicans* strains.

Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 392; DB 21; Length 392;
 Best Local Similarity 100.0%; Pred. No. 1.5e-70;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgggtgttgacactatttgggtgcagaactgatacgaataatgagttattatga 60
 |||||||
 DB 1 tgggtgttgacactatttgggtgcagaactgatacgaataatgagttattatga 60
 QY 61 gaatgaaattttccatcacacatcagtgatgacgaactaaactatttggagt 120
 |||||||
 DB 61 gaatgaaattttccatcacacatcagtgatgacgaactaaactatttggagt 120
 QY 121 ataataagggtatgaataataacatccagaataatacaagagatagaaggagagt 180
 |||||||
 DB 121 ataataagggtatgaataataacatccagaataatacaagagatagaaggagagt 180
 QY 181 taatataatcttggatgaataataactcggttctaattcactatacacactagaagtg 240
 |||||||
 DB 181 taatataatcttggatgaataataactcggttctaattcactatacacactagaagtg 240
 QY 241 aaagctcaatctcaggaagaagaatttatcttcacagattagaagtcagtagtat 300
 |||||||

DB 241 aaagctcaatctcaggaagaagaatttatcttcacagattagaagtcagtagtat 300
 QY 301 aatcaattcgtcccaaatatgagtgatataatcagagccccaattgtattatgatt 360
 |||||||
 DB 301 aatcaattcgtcccaaatatgagtgatataatcagagccccaattgtattatgatt 360
 QY 361 gatagttcgaagttggaaggtacagaatttc 392
 |||||||
 DB 361 gatagttcgaagttggaaggtacagaatttc 392

RESULT 3
 AAA57923
 ID AAA57923 standard; DNA; 392 BP.
 XX
 AC AAA57923;
 XX
 DT 10-OCT-2000 (first entry)
 XX
 DE *Candida albicans* strain SGY-1 TCa2 retrotransposon 5' region.
 XX
 KW Retrotransposon; pCal; TCa2; TY1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 XX
 OS *Candida albicans* strain SGY-1.
 XX
 PN WO200026397-A1.
 XX
 PD 11-MAY-2000.
 XX
 PF 01-NOV-1999; 99WO-NZ00179.
 XX
 PR 30-OCT-1998; 98CA-2249046.
 PR 30-OCT-1998; 98US-0106342.
 XX
 PA (JANC) JANSSEN PHARM NV.
 PI Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;
 DR WPI; 2000-365640/31.
 XX
 XX Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of *Candida* in a sample -
 XX
 PS Example 9; Fig 9; 204pp; English.

The invention relates to novel retrotransposons from the yeast *Candida albicans* which have a copy number of 40-150, preferably 50-100 copies per genome. In particular, the invention relates to the novel *C. albicans* TY1/copia retrotransposon pCal (AA57920), and to the integrated form of this retrotransposon, designated TCa2, and to the novel *C. albicans* retrotransposons 1-28. pCal was initially isolated from *C. albicans* HOG1042 and has a copy number of 50-100 copies per cell. It comprises identical 280 bp long terminal repeats (LTRs) and two open reading frames (ORFs). The first ORF encodes a gag (group antigen) protein, and the second ORF encodes a polyprotein (pol) consisting of an aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and pol ORFs of pCal are in the same reading frame, separated only by a termination codon (TGA). Translation of the pol ORF occurs through the occasional readthrough suppression of the stop codon, which is mediated by the formation of a pseudoknot within the gag-pol mRNA. The retrotransposons of the invention can be used as vectors for in vitro or in vivo transformation and expression. They can thus be used for the delivery and expression of a therapeutic, immunological or immunogenic molecule (e.g., an antigen) and may also be used for eliciting an immunological response in a host organism. They are therefore useful in genetic vaccine compositions and for gene therapy, particularly where the use of retroviral vectors is unsafe or undesirable. Additionally, the retrotransposons may be used to generate

transgenic animals, to detect the presence of *Candida* in a sample, to detect and disrupt genes, and to assign functions to nucleotide sequences. Sequences AAA57923-A57936 represents Tca2 retrotransposon 5' regions from a variety of *C. albicans* strains.

Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
Best Local Similarity 99.5%; Pred. No. 6.9e-70;
Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tttgtgttgacactatttggcgaagaactgatacgaagaatgattattatga 60
DB 1 tttgtgttgacactatttggcgaagaactgatacgaagaatgattattatga 60
QY 61 gaatgaaattttccatcacacatcaggtgatacgaagaactaactattgtgtagt 120
DB 61 gaatgaaattttccatcacacatcaggtgatacgaagaactaactattgtgtagt 120
QY 121 ataaataagggtatgaatataccacacatccagaatatacagaatagaaggagagt 180
DB 121 ataaataagggtatgaatataccacacatccagaatatacagaatagaaggagagt 180
QY 181 tcaatataatcttggatataataactcgttcttaattcactatacacaactagacgtgt 240
DB 181 tcaatataatcttggatataataactcgttcttaattcactatacacaactagacgtgt 240
QY 241 aacgcctcaatcctcaggtgaagaagtattatcattcagattgaagtcagatgagt 300
DB 241 aacgcctcaatcctcaggtgaagaagtattatcattcagattgaagtcagatgagt 300
QY 301 aatcattcgtcccaaatagcgttgaataatcagtcctcagaattgtattatgatt 360
DB 301 aatcattcgtcccaaatagcgttgaataatcagtcctcagaattgtattatgatt 360
QY 361 gatagttcgaagtttgaaggtacagaatttc 392
DB 361 gatagttcgaagtttgaaggtacagaatttc 392

RESULT 4

AAA57927
ID AAA57927 standard; DNA; 392 BP.

AC AAA57927;

PT 10-OCT-2000 (first entry)

DE *Candida albicans* strain ATC-1 Tca2 retrotransposon 5' region.

XX Retrotransposon; pCal; Tca2; Tyl; copia; long terminal repeat; LTR;
KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
KW stop codon suppression; gene delivery; gene therapy vector;
KW genetic vaccine composition; immunogenic; transgenic animal; ds.

XX *Candida albicans* strain ATC-1.

OS WO200026397-A1.

PN 11-MAY-2000.

PF 01-NOV-1999; 99WO-NZ00179.

PR 30-OCT-1998; 98CA-2249046.

PR 30-OCT-1998; 98US-0106342.

XX (JANNC) JANSSEN PHARM NV.

XX Luyten WHM, De Backer MD, Nelissen BJM, Poultier RTM;

DR WPI; 2000-365640/31.

XX Novel retrotransposon expression vectors useful for expressing an
PT antigen, epitope or therapeutic agent, or detecting genes or the
PT presence of *Candida* in a sample -

PS Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*
CC *albicans* which have a copy number of 40-150, preferably 50-100 copies
CC per genome. In particular, the invention relates to the novel *C.*
CC *albicans* Tyl/copia retrotransposon pCal (AAA57920), and to the integrated
CC form of this retrotransposon, designated Tca2, and to the novel *C.*
CC *albicans* retrotransposons 1-28. pCal was initially isolated from *C.*
CC *albicans* H031042 and has a copy number of 50-100 copies per cell. It
CC comprises identical 280 bp long terminal repeats (LTRs) and two open
CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
CC gag and pol ORFs of pCal are in the same reading frame, separated only by
CC a termination codon (TGA). Translation of the pol ORF occurs through
CC the occasional readthrough suppression of the stop codon, which is
CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
CC The retrotransposons of the invention can be used as vectors for *in*
CC *vitro* or *in vivo* transformation and expression. They can thus be used
CC for the delivery and expression of a therapeutic, immunological or
CC immunogenic molecule (e.g., an antigen) and may also be used for
CC eliciting an immunological response in a host organism. They are
CC therefore useful in genetic vaccine compositions and for gene therapy,
CC particularly where the use of retroviral vectors is unsafe or
CC undesirable. Additionally, the retrotransposons may be used to generate
CC transgenic animals, to detect the presence of *Candida* in a sample, to
CC detect and disrupt genes, and to assign functions to nucleotide
CC sequences. Sequences AAA57923-A57936 represents Tca2 retrotransposon 5'
CC regions from a variety of *C. albicans* strains.

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
Best Local Similarity 99.5%; Pred. No. 6.9e-70;
Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tttgtgttgacactatttggcgaagaactgatacgaagaatgattattatga 60
DB 1 tttgtgttgacactatttggcgaagaactgatacgaagaatgattattatga 60
QY 61 gaatgaaattttccatcacacatcaggtgatacgaagaactaactattgtgtagt 120
DB 61 gaatgaaattttccatcacacatcaggtgatacgaagaactaactattgtgtagt 120
QY 121 ataaataagggtatgaatataccacacatccagaatatacagaatagaaggagagt 180
DB 121 ataaataagggtatgaatataccacacatccagaatatacagaatagaaggagagt 180
QY 181 tcaatataatcttggatataataactcgttcttaattcactatacacaactagacgtgt 240
DB 181 tcaatataatcttggatataataactcgttcttaattcactatacacaactagacgtgt 240
QY 241 aacgcctcaatcctcaggtgaagaagtattatcattcagattgaagtcagatgagt 300
DB 241 aacgcctcaatcctcaggtgaagaagtattatcattcagattgaagtcagatgagt 300
QY 301 aatcattcgtcccaaatagcgttgaataatcagtcctcagaattgtattatgatt 360
DB 301 aatcattcgtcccaaatagcgttgaataatcagtcctcagaattgtattatgatt 360
QY 361 gatagttcgaagtttgaaggtacagaatttc 392
DB 361 gatagttcgaagtttgaaggtacagaatttc 392

RESULT 5

AAA57928

ID AAA57928 standard; DNA; 392 BP.
 AC AAA57928;
 XX
 XX 10-OCT-2000 (first entry)
 DE Candida albicans strain ATC-2 Tca2 retrotransposon 5' region.
 XX
 XX Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 OS
 XX Candida albicans strain ATC-2.
 XX
 XX W0200026397-A1.
 XX
 XX 11-MAY-2000.
 XX
 XX 01-NOV-1999; 99WO-NZ00179.
 XX
 XX 30-OCT-1998; 98CA-2249046.
 XX 30-OCT-1998; 98US-0106342.
 XX
 XX (JANC) JANSSEN PHARM NV.
 XX
 XX Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;
 XX
 XX WPI: 2000-365640/31.
 XX
 XX Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -
 XX
 XX Example 9; Fig 9; 204pp; English.
 PS
 XX The invention relates to novel retrotransposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans Ty1/copia retrotransposon pCal (AAA57920), and to the integrated
 CC form of this retrotransposon, designated Tca2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans hog1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AAA57923-457936 represents Tca2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.
 XX
 XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
 Best Local Similarity 99.5%; Pred. No. 6.9e-70;
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 tttgttctgtgcactatttctgtcagaactgacatgaaatgattattatga 60

Db
 1 tttgttctgtgcactatttctgtcagaactgacatgaaatgattattatga 60
 QY 61 gaatgaaaatttttccatcacatcagtgatgacagaactaaactattgtgtgt 120
 Db 61 gaatgaaaatttttccatcacatcagtgatgacagaactaaactattgtgtgt 120
 QY 121 ataatgaagggtatgaataccacatccagaataccaagatagaaggaggtt 180
 Db 121 ataatgaagggtatgaataccacatccagaataccaagatagaaggaggtt 180
 QY 181 tcaatatatctcttgaataaacttgcgttcaatcaactatcacacaactagagtt 240
 Db 181 tcaatatatctcttgaataaacttgcgttcaatcaactatcacacaactagagtt 240
 QY 241 acagcgtcaatctcagtgaaagaatttatatccatagattagaagtcagttgt 300
 Db 241 acagcgtcaatctcagtgaaagaatttatatccatagattagaagtcagttgt 300
 QY 301 aatcattctgcccaaatagtcgttcatataatccagtcctcagatttcatattgatt 360
 Db 301 aatcattctgcccaaatagtcgttcatataatccagtcctcagatttcatattgatt 360
 QY 361 gatagttcgaagtttgaaggtacagaatttc 392
 Db 361 gatagttcgaagtttgaaggtacagaatttc 392

RESULT 6

AAA57931
 ID AAA57931 standard; DNA; 392 BP.
 AC AAA57931;
 XX
 XX 10-OCT-2000 (first entry)
 DE Candida albicans strain F16-1 Tca2 retrotransposon 5' region.
 XX
 XX Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 XX
 XX Candida albicans strain F16-1.
 OS
 XX W0200026397-A1.
 XX
 XX 11-MAY-2000.
 XX
 XX 01-NOV-1999; 99WO-NZ00179.
 XX
 XX 30-OCT-1998; 98CA-2249046.
 XX 30-OCT-1998; 98US-0106342.
 XX
 XX (JANC) JANSSEN PHARM NV.
 XX
 XX Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;
 XX
 XX WPI: 2000-365640/31.
 XX
 XX Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -
 XX
 XX Example 9; Fig 9; 204pp; English.

The invention relates to novel retrotransposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans Ty1/copia retrotransposon pCal (AAA57920), and to the integrated
 CC form of this retrotransposon, designated Tca2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.

CC albicans hOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polypeptide (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pcal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AA57923-AA57936 represents TCA2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.

XX
 SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
 Best Local Similarity 99.5%; Pred. No. 6.9e-70;
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tttgtgttgcactatttgcgcgaacgcatcaatgaagaatgattatata 60
 DB 1 tttgtgttgcactatttgcgcgaacgcatcaatgaagaatgattatata 60
 QY 61 gaatgaaattttccatcacatcagtgatgacagactaatatattgtagt 120
 DB 61 gaatgaaattttccatcacatcagtgatgacagactaatatattgtagt 120
 QY 121 ataataaggatgaataccacatccagataatcaagagatagaaggagagt 180
 DB 121 ataataaggatgaataccacatccagataatcaagagatagaaggagagt 180
 QY 121 ataataaggatgaataccacatccagataatcaagagatagaaggagagt 180
 DB 121 ataataaggatgaataccacatccagataatcaagagatagaaggagagt 180
 QY 181 tcaatataatcttgcgaataaacttcgttcaatcaactaacaactgaagtg 240
 DB 181 tcaatataatcttgcgaataaacttcgttcaatcaactaacaactgaagtg 240
 QY 241 acacgctcaatcgcgaataaagtttatattcattcagattagaagtcagtg 300
 DB 241 acacgctcaatcgcgaataaagtttatattcattcagattagaagtcagtg 300
 QY 301 aatcaatttcgccaataatgagtgatataaattcgcctcagattgtattgt 360
 DB 301 aatcaatttcgccaataatgagtgatataaattcgcctcagattgtattgt 360
 QY 361 gatagttcgaagttgaaggtacgaatttc 392
 DB 361 gatagttcgaagttgaaggtacgaatttc 392

RESULT 7
 ID AA57932
 XX AA57932 standard; DNA; 392 BP.
 AC AA57932;
 XX
 DE 10-OCT-2000 (first entry)
 XX
 XX Candida albicans strain Fl6-2 TCA2 retrotransposon 5' region.
 KW retrotransposon; pCal; TCA2; TY1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polypeptide; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.

XX
 OS Candida albicans strain Fl6-2.
 XX
 PN WO200026397-A1.
 XX
 PD 11-MAY-2000.
 XX
 XX 01-NOV-1999; 99WO-NZ00179.
 XX
 PR 30-OCT-1998; 98CA-2249046.
 PR 30-OCT-1998; 98US-0106342.
 XX
 PA (JANC) JANSSEN PHARM NV.
 XX
 PI Luyten WHM, De Backer MD, Nelissen BJM, Poultier RTM.
 DR WPI; 2000-365640/31.
 XX
 PT Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of candida in a sample
 XX
 PS Example 9; Fig 9; 204pp; English.

CC The invention relates to novel retrotransposons from the yeast Candida
 CC albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular the invention relates to the novel C.
 CC albicans TY1/copia retrotransposon pCal (AA57920), and to the integrated
 CC form of this retrotransposon, designated TCA2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans hOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polypeptide (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AA57923-AA57936 represents TCA2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.

SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 other;

Query Match 6.1%; Score 388.8; DB 21; Length 392;
 Best Local Similarity 99.5%; Pred. No. 6.9e-70;
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tttgtgttgcactatttgcgcgaacgcatcaatgaagaatgattatata 60
 DB 1 tttgtgttgcactatttgcgcgaacgcatcaatgaagaatgattatata 60
 QY 61 gaatgaaattttccatcacatcagtgatgacagactaatatattgtagt 120
 DB 61 gaatgaaattttccatcacatcagtgatgacagactaatatattgtagt 120
 QY 121 ataataaggatgaataccacatccagataatcaagagatagaaggagagt 180
 DB 121 ataataaggatgaataccacatccagataatcaagagatagaaggagagt 180
 QY 181 tcaatataatcttgcgaataaacttcgttcaatcaactaacaactgaagtg 240
 DB 181 tcaatataatcttgcgaataaacttcgttcaatcaactaacaactgaagtg 240

Query Match	Best Local Similarity	Score	DB	Length
Matches 389; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	99.2%; Pred. No. 1.5e-69;	387.2;	21;	392;
Qy 1	1 tgttggttgagcaactatttggtgcagaactgatacgaatgataatgattattatga	60		
Dy 1	1 tgttggttgagcaactatttggtgcagaactgatacgaatgataatgattattatga	60		
Qy 61	gaatgaaaattttccatccacacatccaggtgagtacagaaactaaactattggtagt	120		
Dy 61	gaatgaaaattttccatccacacatccaggtgagtacagaaactaaactattggtagt	120		
Qy 121	ataaataagggtatgaaataccacacatccagaataatacagagataagaggagatt	180		
Dy 121	ataaataagggtatgaaataccacacatccagaataatacagagataagaggagatt	180		
Qy 181	tcaatatatacttggtgataataacttcgtttcaattcaacatacacacacagagcgt	240		
Dy 181	tcaatatatacttggtgataataacttcgtttcaattcaacatacacacacagagcgt	240		
Qy 241	aacagctcaatctccaggtlaaagaaagtttatattccatcagattagaagtcagatg	300		
Dy 241	aacagctcaatctccaggtlaaagaaagtttatattccatcagattagaagtcagatg	300		
Qy 301	aatcatttcgcccccaaatgaagcgtgtataattcagtcctcagattggtattatgatt	360		
Dy 301	aatcatttcgcccccaaatgaagcgtgtataattcagtcctcagattggtattatgatt	360		
Qy 361	gatagttcgaagtttgaaggtacagaatttc 392			
Dy 361	gatagttcgaagtttgaaggtacagaatttc 392			
Result 9	AAA57929	standard; DNA; 392 BP.		
XX	AAA57929;			
XX	10-OCT-2000 (first entry)			
XX	Candida albicans strain SA4-1 TCa2 retrotransposon 5' region.			
XX	Retrotansposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;			
XX	gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;			
XX	Reverse transcriptase; RNaseH; pseudonkt; readthrough translation;			
XX	stop codon suppression; gene delivery; gene therapy vector;			
XX	genetic vaccine composition; immunogenic; transgenic animal; ds.			
XX	Candida albicans strain SA4-1.			
XX	MO200026397-A1.			
XX	11-MAY-2000.			
XX	01-NOV-1999; 99MO-NZ00179.			
XX	30-OCT-1998; 98CA-2249046.			
XX	30-OCT-1998; 98US-0106342.			
XX	(JAMC) JANSSEN PHARM NV.			

Best Local Similarity 99.2%; Pred. No. 1.5e-69; Matches 389; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tttgtgttgcacacttttgcagaaactatcaatgaataatgattatata 60
 Db 1 tttgtgttgcacacttttgcagaaactatcaatgaataatgattatata 60
 QY 61 gaatgaaatatttccatcacacatcaggtgtagacagaactaataatattgtagt 120
 Db 61 gaatgaaatatttccatcacacatcaggtgtagacagaactaataatattgtagt 120
 QY 121 ataataagggtatgaataatcacaatccagaatatccaagagatagaaggagagtt 180
 Db 121 ataataagggtatgaataatcacaatccagaatatccaagagatagaaggagagtt 180
 QY 181 tcaatataatctgtgaataatcacttcgttcaattcactataccaactagacgtgt 240
 Db 181 tcaatataatctgtgaataatcacttcgttcaattcactataccaactagacgtgt 240
 QY 241 acacgctcaatcctcaggtgaagaagtattatcattcattcagattagaagtcgaatgtat 300
 Db 241 acacgctcaatcctcaggtgaagaagtattatcattcattcagattagaagtcgaatgtat 300
 QY 301 aatcattcgtcccaaatagcgtgtataaatcagtcctcagattgtatattgatt 360
 Db 301 aatcattcgtcccaaatagcgtgtataaatcagtcctcagattgtatattgatt 360
 QY 361 gatagttcgaagttgaagttacagaatttc 392
 Db 361 gatagttcgaagttgaagttacagaatttc 392

RESULT 11

AAAS7930
 ID AAAS7930 standard; DNA; 392 BP.

AC AAAS7930;

DT 10-OCT-2000 (first entry)

XX Candida albicans strain SA4-2 Tca2 retrotransposon 5' region.

XX Retrotransposon; pCal; Tca2; Tyl; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.

XX Candida albicans strain SA4-2.

XX WO200026397-A1.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-NZ00179.

XX 30-OCT-1998; 98CA-2249046.

XX 30-OCT-1998; 98US-0106342.

XX (JAMC) JANSSEN PHARM NV.

XX Luyten WHML, De Backer MD, Nelissen BJM, Foulter RTM;

XX WPI; 2000-365640/31.

XX Novel retrotransposon expression vectors useful for expressing an

XX antigen, epitope or therapeutic agent, or detecting genes or the

XX presence of Candida in a sample -

XX Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast Candida

XX albicans which have a copy number of 40-150, preferably 50-100 copies

CC per genome. In particular, the invention relates to the novel C.
 CC albicans Tyl1/copia retrotransposon pCal (AAAS7920), and to the integrated
 CC form of this retrotransposon, designated Tca2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans HOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transfection and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AAAS7923-A57936 represents Tca2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.

XX Sequence 392 BP; 135 A; 58 C; 72 G; 127 T; 0 other;

XX Query Match 6.0%; Score 385.6; DB 21; Length 392;

XX Best Local Similarity 99.0%; Pred. No. 3.1e-69; Matches 388; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tttgtgttgcacacttttgcagaaactatcaatgaataatgattatata 60
 Db 1 tttgtgttgcacacttttgcagaaactatcaatgaataatgattatata 60
 QY 61 gaatgaaatatttccatcacacatcaggtgtagacagaactaataatattgtagt 120
 Db 61 gaatgaaatatttccatcacacatcaggtgtagacagaactaataatattgtagt 120
 QY 121 ataataagggtatgaataatcacaatccagaatatccaagagatagaaggagagtt 180
 Db 121 ataataagggtatgaataatcacaatccagaatatccaagagatagaaggagagtt 180
 QY 181 tcaatataatctgtgaataatcacttcgttcaattcactataccaactagacgtgt 240
 Db 181 tcaatataatctgtgaataatcacttcgttcaattcactataccaactagacgtgt 240
 QY 241 acacgctcaatcctcaggtgaagaagtattatcattcattcagattagaagtcgaatgtat 300
 Db 241 acacgctcaatcctcaggtgaagaagtattatcattcattcagattagaagtcgaatgtat 300
 QY 301 aatcattcgtcccaaatagcgtgtataaatcagtcctcagattgtatattgatt 360
 Db 301 aatcattcgtcccaaatagcgtgtataaatcagtcctcagattgtatattgatt 360
 QY 361 gatagttcgaagttgaagttacagaatttc 392
 Db 361 gatagttcgaagttgaagttacagaatttc 392

RESULT 12

AAAS7934
 ID AAAS7934 standard; DNA; 392 BP.

AC AAAS7934;

DT 10-OCT-2000 (first entry)

XX Candida albicans strain 759-2 Tca2 retrotransposon 5' region.

XX Retrotransposon; pCal; Tca2; Tyl; copia; long terminal repeat; LTR;

KM gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KM reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KM stop codon suppression; gene delivery; gene therapy vector;
 KM genetic vaccine composition; immunogenic; transgenic animal; ds.
 OS Candida albicans strain 759-2.
 XX WO200026397-A1.
 XX 11-MAY-2000.
 XX 01-NOV-1999; 99WO-NZ00179.
 XX 30-OCT-1998; 98CA-2249046.
 PR 30-OCT-1998; 98US-0106342.
 XX (JANC) JANSSEN PHARM NV.
 PA Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;
 DR WPI: 2000-365640/31.
 XX Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -
 XX Example 9; Fig 9; 204pp; English.
 PS The invention relates to novel retrotransposons from the yeast Candida
 XX albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans Ty1/copia retrotransposon pCal (AA57920), and to the integrated
 CC form of this retrotransposon, designated TCa2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans HOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
 CC The retrotransposons of the invention can be used as vectors for in
 CC vitro or in vivo transformation and expression. They can thus be used
 CC for the delivery and expression of a therapeutic, immunological or
 CC immunogenic molecule (e.g., an antigen) and may also be used for
 CC eliciting an immunological response in a host organism. They are
 CC therefore useful in genetic vaccine compositions and for gene therapy,
 CC particularly where the use of retroviral vectors is unsafe or
 CC undesirable. Additionally, the retrotransposons may be used to generate
 CC transgenic animals, to detect the presence of Candida in a sample, to
 CC detect and disrupt genes, and to assign functions to nucleotide
 CC sequences. Sequences AA57923-AA57936 represents TCa2 retrotransposon 5'
 CC regions from a variety of C. albicans strains.
 XX
 XX Sequence 392 BP; 138 A; 59 C; 71 G; 124 T; 0 other;
 SQ
 Query Match 6.0%; Score 384; DB 21; Length 392;
 Best Local Similarity 98.7%; Pred No. 6.5e-69;
 Matches 387; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 tgttggttgacattttgttcagaactgacatgaagaatgagttattatga 60
 Db 1 tgttggttgacattttgttcagaactgacatgaagaatgagttattatga 60
 QY 61 gaatgaaaattttccatcacacatcagtgatgacagaactaatactattgtgtatg 120
 Db 61 gaatgaaaattttccatcacacatcagtgatgacagaactaatactattgtgtatg 120
 QY 121 ataataaaggtatgaataccacacatccagaataatcaacagatagaaggagaggt 180
 Db 121 ataataaaggtatgaataccacacatccagaataatcaacagatagaaggagaggt 180

Db 121 ataataaaggtatgaataccacacatccagaataatcaacagatagaaggagaggt 180
 QY 181 tcaatataatcttcttgataataaacttcgttctaattcaatatacacacaactagacgtgt 240
 Db 181 tcaatataatcttcttgataataaacttcgttctaattcaatatacacacaactagacgtgt 240
 QY 241 aacagctcaatttcagtgataaagaagttatattccatcacagattagaagtcagatgcat 300
 Db 241 aacagctcaatttcagtgataaagaagttatattccatcacagattagaagtcagatgcat 300
 QY 301 aatcatttcgtcccaattagcgttctgataaattcagtcctccagaattgtattatgatt 360
 Db 301 aacatttcgtcccaattagcgttctgataaattcagtcctccagaattgtattatgatt 360
 QY 361 gatagtttcgaagtttgaaggtacagaatttc 392
 Db 361 gatagtttcgaagtttgaaggtacagaatttc 392
 RESULT 13
 AA57935
 ID AA57935 standard; DNA: 392 BP.
 AC AA57935;
 XX 10-OCT-2000 (first entry)
 DT
 DE Candida albicans strain p30 TCa2 retrotransposon 5' region.
 XX Retrotransposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
 KW stop codon suppression; gene delivery; gene therapy vector;
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.
 XX
 OS Candida albicans strain p30.
 XX WO200026397-A1.
 XX 11-MAY-2000.
 XX 01-NOV-1999; 99WO-NZ00179.
 XX 30-OCT-1998; 98CA-2249046.
 PR 30-OCT-1998; 98US-0106342.
 XX (JANC) JANSSEN PHARM NV.
 PA Luyten WHML, De Backer MD, Nelissen BJM, Poulter RTM;
 PT WPI: 2000-365640/31.
 DR Novel retrotransposon expression vectors useful for expressing an
 PT antigen, epitope or therapeutic agent, or detecting genes or the
 PT presence of Candida in a sample -
 XX Example 9; Fig 9; 204pp; English.
 PS The invention relates to novel retrotransposons from the yeast Candida
 XX albicans which have a copy number of 40-150, preferably 50-100 copies
 CC per genome. In particular, the invention relates to the novel C.
 CC albicans Ty1/copia retrotransposon pCal (AA57920), and to the integrated
 CC form of this retrotransposon, designated TCa2, and to the novel C.
 CC albicans retrotransposons 1-28. pCal was initially isolated from C.
 CC albicans HOG1042 and has a copy number of 50-100 copies per cell. It
 CC comprises identical 280 bp long terminal repeats (LTRs) and two open
 CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
 CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
 CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
 CC gag and pol ORFs of pCal are in the same reading frame, separated only by
 CC a termination codon (TGA). Translation of the pol ORF occurs through
 CC the occasional readthrough suppression of the stop codon, which is
 CC mediated by the formation of a pseudoknot within the gag-pol mRNA.

CC The retrotransposons of the invention can be used as vectors for in
CC vitro or in vivo transformation and expression. They can thus be used
CC for the delivery and expression of a therapeutic, immunological or
CC immunogenic molecule (e.g., an antigen) and may also be used for
CC eliciting an immunological response in a host organism. They are
CC therefore useful in genetic vaccine compositions and for gene therapy,
CC particularly where the use of retroviral vectors is unsafe or
CC undesirable. Additionally, the retrotransposons may be used to generate
CC transgenic animals, to detect the presence of *Candida* in a sample, to
CC detect and disrupt genes, and to assign functions to nucleotide
CC sequences. Sequences AA57923-AS7936 represents TCA2 retrotransposon 5'
CC regions from a variety of *C. albicans* strains.

XX Sequence 392 BP; 139 A; 55 C; 71 G; 127 T; 0 other;

Query Match 6.0%; Score 384; DB 21; Length 392;
Best Local Similarity 98.7%; Pred. No. 6,5e-69;
Matches 387; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tttgtgttgcacattttgtgtcagaacatgcatgaataatgattatga 60
DB 1 tttgtgttgcacattttgtgtcagaacatgcatgaataatgattatga 60
QY 61 gaatggaataatcttcacatcacatcagtgatgacagaactaactatattgtagt 120
DB 61 gaatggaataatcttcacatcacatcagtgatgacagaactaactatattgtagt 120
QY 121 ataataagggtatgaataaccacatcccgaaatcacaagatgataaggagagagt 180
DB 121 ataataagggtatgaataaccacatcccgaaatcacaagatgataaggagagagt 180
QY 181 tcaatataatctgtgataataactcgttcaattcattcacaacacataagctgt 240
DB 181 tcaatataatctgtgataataactcgttcaattcattcacaacacataagctgt 240
QY 241 acacgctcaatcagtgaaagaagttatattccatcagatagaagatgagat 300
DB 241 acacgctcaatcagtgaaagaagttatattccatcagatagaagatgagat 300
QY 301 aatcattcgtcccaaatcagtgatgataaattcagtcctcagaattgtattgatt 360
DB 301 aatcattcgtcccaaatcagtgatgataaattcagtcctcagaattgtattgatt 360
QY 361 gatagttcgaagttgaagttacaagaatttc 392
DB 361 gatagttcgaagttgaagttacaagaatttc 392

SURF 14

AA57924 standard; DNA; 392 BP.

AA57924;

10-OCT-2000 (first entry)

Candida albicans strain SGV-2 TCA2 retrotransposon 5' region.

Retroransposon; pCal; TCA2; TY1; copia; long terminal repeat; LTR;
gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
stop codon suppression; gene delivery; gene therapy vector;
genetic vaccine composition; immunogenic; transgenic animal; ds.

Candida albicans strain SGV-2.

MO200026397-A1.

11-MAY-2000.

01-NOV-1999; 99MO-NZ00179.

XX

PR 30-OCT-1998; 98CA-2249046.
PR 30-OCT-1998; 98US-0106342.

XX (JANC) JANSSEN PHARM NV.

PI Luyten WHML, De Backer MD, Nelissen BJM, Poulier RTM;
XX WPI: 2000-365640/31.

PT Novel retrotransposon expression vectors useful for expressing an
PT antigen, epitope or therapeutic agent, or detecting genes or the
PT presence of *Candida* in a sample.

XX Example 9; Fig 9; 204pp; English.

CC The invention relates to novel retrotransposons from the yeast *Candida*
CC *albicans* which have a copy number of 40-150, preferably 50-100 copies
CC per genome. In particular, the invention relates to the novel *C.*
CC *albicans* TY1/copia retrotransposon pCal (AA57920), and to the integrated
CC form of this retrotransposon, designated TCA2, and to the novel *C.*
CC *albicans* retrotransposons 1-28. pCal was initially isolated from *C.*
CC *albicans* HOG1042, and has a copy number of 50-100 copies per cell. It
CC comprises identical 280 bp long terminal repeats (LTRs) and two open
CC reading frames (ORFs). The first ORF encodes a gag (group antigen)
CC protein, and the second ORF encodes a polyprotein (pol) consisting of an
CC aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The
CC gag and pol ORFs of pCal are in the same reading frame, separated only by
CC a termination codon (TGA). Translation of the pol ORF occurs through
CC the occasional readthrough suppression of the stop codon, which is
CC mediated by the formation of a pseudoknot within the gag-pol mRNA.
CC The retrotransposons of the invention can be used as vectors for in
CC vitro or in vivo transformation and expression. They can thus be used
CC for the delivery and expression of a therapeutic, immunological or
CC immunogenic molecule (e.g., an antigen) and may also be used for
CC eliciting an immunological response in a host organism. They are
CC therefore useful in genetic vaccine compositions and for gene therapy,
CC particularly where the use of retroviral vectors is unsafe or
CC undesirable. Additionally, the retrotransposons may be used to generate
CC transgenic animals, to detect the presence of *Candida* in a sample, to
CC detect and disrupt genes, and to assign functions to nucleotide
CC sequences. Sequences AA57923-AS7936 represents TCA2 retrotransposon 5'
CC regions from a variety of *C. albicans* strains.

SQ Sequence 392 BP; 135 A; 57 C; 74 G; 126 T; 0 other;

Query Match 6.0%; Score 382.4; DB 21; Length 392;
Best Local Similarity 98.5%; Pred. No. 1.4e-68;
Matches 386; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 tttgtgttgcacattttgtgtcagaacatgcatgaataatgattatga 60
DB 1 tttgtgttgcacattttgtgtcagaacatgcatgaataatgattatga 60
QY 61 gaatggaataatcttcacatcacatcagtgatgacagaactaactatattgtagt 120
DB 61 gaatggaataatcttcacatcacatcagtgatgacagaactaactatattgtagt 120
QY 121 ataataagggtatgaataaccacatcccgaaatcacaagatgataaggagagagt 180
DB 121 ataataagggtatgaataaccacatcccgaaatcacaagatgataaggagagagt 180
QY 181 tcaatataatctgtgataataactcgttcaattcattcacaacacataagctgt 240
DB 181 tcaatataatctgtgataataactcgttcaattcattcacaacacataagctgt 240
QY 241 acacgctcaatcagtgaaagaagttatattccatcagatagaagatgagat 300
DB 241 acacgctcaatcagtgaaagaagttatattccatcagatagaagatgagat 300
QY 301 aatcattcgtcccaaatcagtgatgataaattcagtcctcagaattgtattgatt 360
DB 301 aatcattcgtcccaaatcagtgatgataaattcagtcctcagaattgtattgatt 360

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: gb_est110:*
 191: gb_est111:*
 192: gb_est112:*
 193: gb_est113:*
 194: gb_est114:*
 195: gb_est115:*
 196: gb_est116:*
 197: gb_est117:*
 198: gb_est118:*
 199: gb_est119:*
 200: gb_est120:*
 201: gb_est121:*
 202: gb_est122:*
 203: gb_est123:*
 204: gb_est124:*
 205: gb_est125:*
 206: gb_est126:*
 207: gb_est127:*
 208: gb_est128:*
 209: gb_est129:*
 210: gb_est130:*
 211: gb_est131:*
 212: gb_est132:*
 213: gb_est133:*
 214: gb_est134:*
 215: gb_est135:*
 216: gb_est136:*
 217: gb_est137:*
 218: gb_est138:*
 219: gb_est139:*
 220: gb_est140:*
 221: gb_est141:*
 222: gb_est142:*
 223: gb_est143:*
 224: gb_est144:*
 225: gb_est145:*
 226: gb_est146:*
 227: gb_est147:*
 228: gb_est148:*
 229: gb_est149:*
 230: gb_est150:*
 231: gb_est151:*
 232: gb_est152:*
 233: gb_est153:*
 234: gb_est154:*
 235: gb_est155:*
 236: gb_est156:*
 237: gb_est157:*
 238: gb_est158:*
 239: gb_est159:*
 240: gb_est160:*
 241: gb_est161:*
 242: gb_est162:*
 243: gb_est163:*
 244: gb_est164:*
 245: gb_est165:*
 246: gb_est166:*
 247: gb_est167:*
 248: gb_est168:*
 249: gb_est169:*
 250: gb_est170:*
 251: gb_est171:*
 252: gb_est172:*
 253: gb_est173:*
 254: gb_est174:*
 255: gb_est175:*
 256: gb_est176:*
 257: gb_est177:*
 258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	97.8	1.5	569	102	A1813079	A1813079 20E11 Plint
2	85	1.3	905	245	AZ550256	AZ550256 ENTEV58TTC
3	83.4	1.3	1101	219	CNS0039G	AL063921 Drosophill1
4	82.4	1.3	849	245	AZ546009	AZ546009 ENTFW53TTC
5	81.8	1.3	1101	219	CNS001LW2	AL078714 Drosophill1
6	80.6	1.3	641	236	AQ946120	AQ946120 Sheared L
7	80	1.2	877	245	AZ531291	AZ531291 ENTBQ34TTC
8	79.4	1.2	749	238	AQ129392	AZ129392 OSJNB01C
9	79	1.2	657	224	AQ156164	AQ156164 nbxb00024TTC
10	78.2	1.2	900	245	AZ549980	AZ549980 ENTPDP94TTC
11	77.8	1.2	912	245	AZ551092	AZ551092 ENTFJ722TTC
12	77.6	1.2	1101	219	CNS0021J	AL061936 Drosophill1
13	77.6	1.2	1101	219	CNS012P8	AL101846 Drosophill1
14	77.4	1.2	617	227	AQ368185	AQ368185 loxb00011H
15	75.4	1.2	908	245	AZ548467	AZ548467 ENTEK30TTC
16	74.4	1.2	660	219	CNS00U07	AL091453 Arabidopses
17	74	1.2	831	234	AQ685895	AQ685895 nbcb0026T
18	74	1.2	974	219	CNS00TTC	AL075432 Drosophill1
19	73.6	1.1	1001	219	CNS0155H	AL105023 Drosophill1
20	72.2	1.1	1101	219	CNS01V72	AL108536 Drosophill1
21	71.2	1.1	843	245	AZ551618	AZ551618 ENTPDV54TTC
22	71	1.1	1147	219	CNS003B0	AL0644078 Drosophill1
23	70.6	1.1	1447	80	BF273312	BF273312 GA_EB001
24	70.4	1.1	605	227	AQ367630	AQ367630 loxb00002T
25	70.4	1.1	700	235	AQ940248	AQ940248 Sheared L
26	70	1.1	891	248	AZ683582	AZ683582 ENTKK47TTC
27	70	1.1	963	219	CNS00U4L	AL104918 Drosophill1
28	69.6	1.1	1025	219	CNS014J2	AL104241 Drosophill1
29	69.4	1.1	739	174	BG127728	BG127728 EST473374
30	69.2	1.1	813	219	CNS0119M	AL093988 Drosophill1
31	69.2	1.1	910	219	CNS0113P	AL092775 Drosophill1
32	69	1.1	867	219	CNS0054A	AL0577618 Drosophill1
33	68.6	1.1	561	230	AQ577014	AQ577014 nbxb009000
34	68.6	1.1	610	230	AQ576975	AQ576975 nbxb0090R
35	68.6	1.1	649	116	AM496652	AM496652 KP0408_Y
36	68.4	1.1	768	234	AQ870251	AQ870251 nbcb0036B5
37	67.8	1.1	536	258	TA287H07Q	TA287H07Q
38	67.8	1.1	1038	219	CNS0108N	AL048498 T_buceai
39	67.4	1.0	437	221	CNS030FO	AL098657 Drosophill1
40	67.2	1.0	1101	219	CNS00EVL	AL253265 Tetraodon
41	67	1.0	951	227	AZ676519	AL069706 Drosophill1
42	67	1.0	994	241	ENMGV51TTC	AZ676519 ENMGV51TTC
43	66.8	1.0	956	219	CNS04NOV	AL128972 Tetraodon
44	66.4	1.0	478	219	CNS014V0	AL104646 Drosophill1
45	66.4	1.0	865	152	BG309498	AL091029 Arabidopses
						BG309498 HVMSEC001

ALIGNMENTS

Accession	Description
AL8183079	20E011.P14
A2550256	ENTVE58T7
AL0063921	Drosophila1
A2546009	ENTVE53T7
AL078714	Drosophila1
A0946120	Sheared.D
A2531291	ENTB034T7
A1283932	OS1NBH01C
A0156164	nbx000022
A2549580	ENTD054T7
A2551092	ENTFJ22T7
AL061936	Drosophila1
AL011846	Drosophila1
A0366185	tox000111
A2546487	ENTB0530T7
AL091453	Arabidops
A0868895	nbx00266
AL075432	Drosophila1
AL015023	Drosophila1
AL108536	Drosophila1
A2551618	ENTDV54T7
AL060478	Drosophila1
B273312	GA_EB01
A0367630	tox000022
A0940248	Sheared.D
A2683582	ENTKR47T7
AL049216	Drosophila1
AL104218	Drosophila1
B0127278	EST413374
AL096988	Drosophila1
AL099775	Drosophila1
A0577014	nbx000900
A0576975	nbx000900
AM496652	K040408.Y
A0870251	nbx00306F
AL088498	T_brucel1
AL098657	Drosophila1
AL253266	Tetradion
AL063706	Drosophila1
A2676915	ENTAGV51T7
A2196972	ENTG001
AL104666	Drosophila1
AL091028	Arabidops
B0309496	HYSMC001

Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
, NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: T3.

FEATURES
source

Seq primer: T3.

1. .569

```
/organism="Pinus taeda"  
/db_xref="taxon:3352"
```

```
/clone_lib="Pine Lambda Zap Xylem library"  
/tissue_type="differentiating xylem"
```

```

Vector: Lambda Zap; Site_1: EcORI; Site_2: XhoI;
Differentiating xylem was collected from the main stem o

```

a 35-year old loblolly pine tree harvested during the growing season. RNA isolation and library preparation followed the methods of Allona et al., PNAS 95:6693-8, 1998"

BASE COUNT	206 a	74 c	120 g	169 t
ORIGIN				

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match
Post Topic

Matches 270; Conservative	0;	Mismatches 287;	Indels	0;	Gaps	0;
Best Local Similarity	48.5%;	Pred. No.	1.3e-10;			

4730 actataagattatgcacatatagttgtgaattagtagatgacaaattcaacatttagac 4789

Db
13 ACAATTAGATTACTGATTGCGCTTAGCCACAACAAATATTTGACACATTCATCAATTGGAT 72

4790 gtcgagtcgagcgtatctaataatgcctctattactcaattcaatccaatttatgtcttcct 484

Db 73 GTCAATCTGCTTTCTAAATGGTATTTAAAGSAGAAGCTTATTTCGTTCAACTGAA 132

QY 4850 cctaatacgtacaccttgaagaanaaccattgttggttatigaacgcttcgtctatgq 490

Db 133 GGGTTTGTCAACAGGGACAGGACATCTGGTTGCAGGTAAAGAAAGCCATTGTATGCT-192

QY 4910 ttaaacagtcaggtttggaatggtatcacactatcaaaagatattgaagacattgct 496

Db 193 TTGAACACGGCACCCAGGTCATGTAATGTAAAGATTGATACCTCTCTTCTCAGATGGA 252

4970 ttactcaagtttaccacaatgaatgatttaccattgaatatgaagaaagatcacta 5021

db 253 TTGTGAAAGCAAAAATGATCTTAAGTGTATGTTAAGAAAGATCAAAAAAGGAATGTT 312

[illegible]

313 GCCTTAATATCGTTTGATGCGGATGCAATTCATTAATTAACAGCGACGTCGCGCAGCAAGCGTAAATTT 3272

5090 gataaattttatcgaatcaatttaagagatcatttttgaatctaaatctttctatataatctatcc 514
 cc 3cctttatattcttttggatgattctcattttatmacagtagtgcctcaccaagttattt 3/2
 212 gctttatattcttttggatgattctcattttatmacagtagtgcctcaccaagttattt 3/2

gacaaatccgcggagcaaacattcaggadgaatcatcacaagaactcaagtcgttcgtcgtaaatatcac 514

373 GAAGAAATTTAAAACCTTTCACGCTTGCTGCACACCGCAATTGGCCATAAATGACAACATTCGCCAT 423

[illegible][illegible]

5310 ++++-----
DB 433 TACTGTTAAAGATTGAGGATGGAGAGAACCTGGTAAACCTTGATCACTCAAGCAA 492

5210 ttctccaagaatttacttaagatcttcaaacacagatgacccataatggyaaaaacataccc 526

Dd 493 TATTCAAAGGAGATTCCATAAAAGTTCATATATGACAGATTGCAAAAGCTATGTCACTCCC 552

QY 5270 tggattccgaatgacaa 5286
| | ||||| ||
FCC

Db 553 TTAGACAGCATGCAAA 569

RESULT 2

AZ550256/c	905 bp	DNA	GSS	14-NOV-2000
LOCUS	AZ550256			

DEFINITION: ENTEV58TR Entamoeba histolytica Sheared DNA Entamoeba histolyt


```

QY 28 gaacatgacatgaatgatgtatattatgaagaatgaaatttccatcacacac 87
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1100 KARRMGDDTMDRDRKDDMTWTKWTKWRADRRRAGADADWADDDAGTWTWTTW 1041
QY 88 aggtgatgacagaactaaactatgtgtgtatataaataaggtatgaataccacat 147
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1040 MWWWATWTDWMDKWWWATWATWATWATWATWATWATWATWATWATWATWATW 981
QY 148 cccagaatataccagatagaaggaaggttccatataatattcttgaataaact 207
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 980 GRDDGKRRKDKDKDKDDDDKGGKKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 921
QY 208 tgcgttcaatcactacacacacacacacacacacacacacacacacacacac 267
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 920 DDDGAGDKDDKGGKADDDDTGTDDDDKDKWMDAKGWTGATWATAATDDWMMG 861
QY 268 ttatattcactcagatagaagtcagatgaatacattcgtcccaatagcgtgt 327
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 860 WADADWMTWDAADWADWADWADWADWADWADWADWADWADWADWADWADWAD 801
QY 328 ataattcactcactcactcactcactcactcactcactcactcactcactcact 387
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 800 RRAADKRRADRDADATWTTTWTTRDTPDKWTKTPTWADRTWDRDDDDDR 741
QY 388 attcacaagatgagtcgcgaagaatgatatacagaagaagatgataagat 447
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 740 AGTAGRRWRTRWRKRRRDRDWDADADDTARDDRRRGGDAGDKGKTKGRRRR 681
QY 448 tgaatcagatcactcactcactcactcactcactcactcactcactcactcact 507
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 680 RATWTDWMDWADWMTWTTTDTDDWDRRRRRKRRRRRRRRRRRRRRRRRRRR 621
QY 508 ttccataaatttaattgaactcactcactcactcactcactcactcactcactcact 567
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 620 KMDWKTARBRWMDWADTWDARADAKARARARARARARARARARARARARAR 561
QY 568 atggaatgagaatataatattcactcactcactcactcactcactcactcact 627
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 560 TWTWTAARAAMWMAWMTTATWTTWTTTWTWTTTWTWTTTWTWTTTWTWTT 501
QY 628 tgaactaactcactcactcactcactcactcactcactcactcactcactcact 687
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 500 AAWTAAAMAAAAAATTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWT 441
QY 688 ttggtgttacaatgtgtt 708
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 440 TTTWTTWMAATTTATTTTWT 420

```

```

RESULT 4
LOCUS A2546009 849 bp DNA GSS 14-NOV-2000
DEFINITION ENFM537F Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION A2546009
VERSION A2546009.1 GI:11167130
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 849)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute of Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: b.loftus@tigr.org

```

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 26
 High quality sequence stop: 796.

FEATURES

source

1..849

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHD31; Site: 1; Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 199 a 236 c 43 g 371 t

ORIGIN

Query Match 1.3%; Score 82.4; DB 245; Length 849;
 Best Local Similarity 46.7%; Pred. No. 2.9e-07;
 Matches 297; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

```

QY 3539 gagagatcaaatgtctgaacagatcattcgtgtcgcgtgttgaagaacacg 3598
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 GAAGCAATGATATAGAAAGATGATGATGATGATGATGATGATGATGATGAT 638
QY 3599 tcagaagatcgcactcctcactcactcactcactcactcactcactcactcact 3658
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 578
QY 3659 cccctgactcgtcactaagaatcgtgtaactgaagaataagatgaagaataat 3718
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 GAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 518
QY 3719 tcattccggggggtgatgatgatcgtcgtgtgtaactccaatgtggaactaat 3778
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 458
QY 3779 gtgaaacagagagatcgtcgtgaacagtcacatccaaatgaagaagttctcaagggaga 3838
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 GAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 398
QY 3839 atacttaatgaacaactgatataagttgtaactcgtgtcctaagttatgaagaataa 3898
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 338
QY 3899 atctctccttaattcaatgaatgatcactcactcactcactcactcactcactcact 3958
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 278
QY 3959 agcaatcacaacagatcgcacatccatcgaagaataatcaccacagttatgaag 4018
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 ---GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 221
QY 4019 aaaaatcgtgaataatccaaaacacatgaagaatcctcgtcgtatagaagatgat 4078
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 161
QY 4079 gaattgaaacgtataatgtgtgaatgagaatgagatgaatgaatgaatgaatgtgt 4138
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 101

```

QY 4139 gaagcaatccaccatgaataaataatgatt 4174
 Db 100 GAAGATGATGATGATGATGATGATGATGAT 65
 RESULT 5
 LOCUS CDS00L72 1101 bp DNA GSS 14-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TERT3 end of BAC: BACR48p19 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL078714 GI:5102004
 VERSION AL078714.1 GI:5102004
 KEYWORDS fruit fly.
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 JOURNAL Direct Submission
 COMMENT Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see: http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 FEATURES
 source Location/Qualifiers
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR48p19"
 /note="end : TERT3"
 BASE COUNT 469 a 6 c 69 g 151 t 406 others
 ORIGIN
 Query Match 1.3%; Score 81.8; DB 219; Length 1101;
 Best Local Similarity 29.2%; Pred. No. 4.1e-07;
 Matches 212; Conservative 166; Mismatches 344; Indels 3; Gaps 1;
 QY 1177 aagcaacaatagaactaattatgtgttaccgcgcgcgagaacaaag 1236
 Db 273 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 332
 QY 1237 aaacccgagagagaccctcgtgacaactcgaagaacctgaatcagaagaag 1296
 Db 333 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 392
 QY 1297 aaataagaacatccaaatcagatacgaataagtgaaaaaagaagaaga 1356
 Db 393 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 452
 QY 1357 aactcactggaatgaagaacagtgctgctcattatgtgtaataatatac 1416
 Db 453 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 512
 QY 1417 ttgagcaaaacacgcttccagtagaagaattcattccttaatgcttctga 1476
 Db 512

Db 513 ADDKMAAMDTTNTTAMTWTWTTATATATTTTTTTTTTTTTTTTAKTKATDWTKKAKTAAT 572
 QY 1477 aatgaatttaaagggttaaggttaa---caaglatcagtgatgatactggtccac 1533
 Db 573 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 632
 QY 1534 aatactgctgtgacaataaagataatgctgaatgtaaggaagcaacaatga 1593
 Db 633 AATTTTWTWTWMAATTTTTTTTTTTTTTTTTTTTTTTTATTAATAAATAAATTTTDTW 692
 QY 1594 ttctgtcgtatggtgctacatcattagaagcagattgattgagatcattcag 1653
 Db 693 AAAMWTTTAKKKKAAADAKKWDAAKWDAAKWDAAKWDAAKWDAAKWDAAKWDAAKWDAAK 752
 QY 1654 cgtatgctcgcattcagatcagtaagaatcattgattatccagaagctcctta 1713
 Db 753 GKGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGK 812
 QY 1714 tctgattgaacaataatgaagaacagattgaatcattatcattactaagaatc 1773
 Db 813 KKAATAKKAADRTTKTKWDAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 872
 QY 1774 gatgtatttaaccaaatatgctcctactatattgcttcaagaagaatgctgta 1833
 Db 873 KDAATAKKAATK 932
 QY 1834 tcttattggtcctcaatcagtaagaatccttgaatgattgattatgattg 1893
 Db 933 GKGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGK 992
 QY 1894 ttgg 1898
 Db 993 KDKK 997
 RESULT 6
 LOCUS A0946120/c 641 bp DNA GSS 27-JAN-2000
 DEFINITION Sheared DNA-46J23.TR Sheared DNA Trypanosoma brucei genomic clone
 ACCESSION A0946120 GI:6769385
 VERSION A0946120.1 GI:6769385
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 REFERENCE 1 (bases 1 to 641)
 AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Villu, E., Melville, S., Donelson, J., Fraser, C. and Adams, M.
 TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
 JOURNAL Unpublished (1999)
 COMMENT Other GSSs: Sheared DNA-46J23.TF
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA;
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tldb/mdb/tbdtb/.
 Seg primer: M13-Reverse
 Class: shotgun.
 FEATURES
 source Location/Qualifiers
 1..641
 /organism="Trypanosoma brucei"
 /strain="TRED927/4 GUTat 10.1"
 /db_xref="taxon:5691"

Query Match	1.38;	Score 80.6;	DB 236;	Length 641;
Best Local Similarity	46.88;	Pred. No. 6.7e-07;		
Matches 254; Conservative	0;	Mismatches 289;	Indels 0;	Gaps 0

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 877)	Loftus,B., Van Aken,S. and Fraser,C.	Determination of clone end sequences from Entamoeba histolytica		
		HM1:IMSS sheared DNA library		
		Unpublished (2000)		
	Contact: Brendan J Loftus			
	Department of Eukaryotic Genomics			
	The Institute for Genomic Research			
	9712 Medical Center Dr., Rockville, MD 20850, USA			
	Tel.: 301 838 0200			
	Fax: 301 838 0208			
	Email: b1loftus@tigr.org			
	Clones are derived from the Entamoeba histolytica HM1:IMSS sheared			
	DNA library			
	Seq primer: M13-Reverse			
	Class: Shotgun			
	High quality sequence start: 22			
	High quality sequence stop: 829.			
	Location/Qualifiers			
	1..877			

[illegible]

```

/strain="HM1:IMSS"
/db xref="taxon:5759"
/clone.lib="Entamoeba histolytica Shared DNA"
/notes="Vector: PHOSI, Site:1, Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.6., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
light size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A practical Approach, eds. M. Vaudin and B.
Barrell, Oxford university Press, 1999)."
```

RESULT	7
AZ531291/c	
LOCUS	AZ531291 877 bp DNA
DEFINITION	Entamoeba histolytica GSS 03-NOV-2000
ACCESSION	ENTB0344R Entamoeba histolytica Sheared DNA
VERSION	AZ531291.1 GI:11085638
SOURCE	GSS.
ORGANISM	Entamoeba histolytica. Entamoeba histolytica. Eukaryota; Entamoebidae; Entamoeba.

Query Match	1.2%	Score 80	DB 245	Length 877
Best Local Similarity	44.4%	Pred No. 9.5e-07		
Matches 320	Conservative	0	Mismatches 400	Indels 0
			Gaps	0
QY 3539	gaggagtcacaaatctcgagcaacagatcatttcggtgcgcgctgtgttaagaacaagtc	3598		
Db 806	GAGCAATTTGTAATTAAAGAAAGACAGAGAGATGATGATGATGAAGAGAAAGATGAT	747		
QY 3599	tccgaagatctgatactctctacaccccaagaactagctgcgaagatggaacagcaaat	3658		
Db 746	GAGGATGATGATGATGAAGAGATGATGATGAAGAGATGATGATGAAGAGATGATGAT	687		
QY 3659	cctcgttactgcgctaataagaatcgggtgaactgggaaaaatagatgaggggaataat	3718		
Db 686	GATCAAGAGAGATGATGAAGAGATGATGAAGAGAGAGAGATGATGATGAAGATGAT	627		
QY 3719	tcaattccgggggtatgatgtatctctgcgcgatccaactcaaatglttgacaatcaat	3778		
Db 626	GAGGACGATGAACACACCAATATGTAATTAGAAAGTGATGATGAAGAGAGAGAT	567		
QY 3779	gttgaacacgagatgctggtgtaacagttcccaattccaagaacgaagttctccaagggaga	3838		
Db 566	GATGATGAAGATGATGAAGACGATGAAGCGACGCAATATGGAATTGAAGATGATGAT	507		
QY 3839	atacttaatggaacaaactatagttgatatactgttctlaaagtattatggaaatgaaaa	3898		
Db 506	GATGAAGAGAGATGATGAAGAGAGAGATGATGATGAAGATGATGAAGATGAAGAA	447		
QY 3899	atctctcctataatcattagatgatcatcactggaactgtgcaacacgcgggaatgat	3958		
Db 446	GAGATGAAGAGATGATGAAGAGATGATGAAGATGATGATGAAGAGAGATGAT	387		

ACCESSION A2549980
 VERSION A2549980.1 GI:11175122
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 900)
 AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HMI:IMSS sheared DNA library
 COMMENT Unpublished (2000)
 CONTACT: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: b1o@tugr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: Shotgun
 High quality sequence start: 20
 High quality sequence stop: 850.
 Location/Qualifiers
 1..900
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"

Query Match	1.2%	Score 78.2;	DB 245;	Length 900;
Best Local Similarity	45.4%;	Pred. No. 2.3e-06;		
Matches 281;	Conservative	0;	Mismatches 338;	Indels 0;
Gaps	0			
3582	ttgttaagaanaacagcttctaagaagttgycataacttcctgactaccgaactagtcgtc	3641		
Db	227	ttgaatttgaagagagacgtttcaaatgtatgaactatgaatttgaaaaacacattgaacatt	286	
Qy	3642	aagatgaacagcaaaaatccctcgtaactgcygctaataagaatccggttaactgaaaaaatag	3701	
Db	287	tgatgatcagcagacgatttgacttgaagaagtgaagaatgacgaatgacgaatgacgaatgaac	346	
Qy	3702	atgaggggagaaaatatcttaattcccgggggggtatgatagatctgctcgatcaactcaa	3761	
Db	347	aagattgaacagcagacgacgatgatgaagacgattgaagatgatgaagatgatgaagatgatg	406	
Qy	3762	atgttgagcaactaactgttgaacacaggagatgctggttaaacagtcacaaatcaagaacgaag	3821	
Db	407	aagatgatcagacatgatgaagatgattgaagaagatgaagaagatgaagaagatgaagacgacg	466	
Qy	3822	ttcttcagaagggaagaacttaattgaacaaactgabatagttgabatctgttgcctaaag	3881	
Db	467	atgaagatgatgaagaatgatgatgaagatgatgaagatgatgaagatgatgaagatgatgaag	526	
Qy	3882	ttattgagatgaaaaaatctctcctaattcaattagatgatgacatcatgaactgtgcta	3941	

QY 1713 ttgtgagtttgaacaaattgaagacgagatttaattcttattactaaagaatcaq 1772

CC TCGTACGCTTATTTAAATAAAGACMAAAATTAAAAAAA 525

QY 1272 agaagaaacctaagaagaagaataaagaacatccaataatcagataacgataaag 1331

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjoftus@tigr.org

Clones are derived from the *Entamoeba histolytica* HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse

Class: shotgun
High quality sequence start: 17
High quality sequence stop: 828.
Location/Qualifiers

FEATURES

1. .908
/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:3759"

/clone_lib="Entamoeba histolytica sheared DNA"

/note="Vector: pHOSt1; Site: 1; Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of *E. histolytica*

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barelli, Oxford University Press, 1999)."

BASE COUNT 434 a 16 c 289 g 169 t
ORIGIN

Query Match 1.28; Score 75.4; DB 245; Length 908;
Best Local Similarity 44.18; Pred. No. 9.3e-06;
Matches 316; Conservative 0; Mismatches 401; Indels 0; Gaps 0;

```
OY 3588 aagaacagctccagaagatgcttacttctgactccagaagatgcttgaagatg 3647
DB 23 AAGAAGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAG 82
OY 3648 aacagcaaatccctcgttactgcttaataagaatcgggtactgtaaaaaa 3707
DB 83 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 142
OY 3708 gagaataattcattccgggggggtgatgatgattcgtcgatgtaactcaat 3767
DB 143 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 202
OY 3768 agcaatcattgtaagaacagagatgctgtaacagtcacatccagaagttctc 3827
DB 203 AAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 262
OY 3828 aagaagggaagaatacttaataaactgatatagttgactgltgtaagttatg 3887
DB 263 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 322
OY 3888 agaataaaaaatctctcatttaattcattagatgataactgtaactgta 3947
DB 323 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 382
OY 3948 cgggaatgatgcaattcaacagatccgacatcattcgaaaaatgaataatcc 4007
DB 383 AAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 442
OY 4008 tgattaatgagaataactgaataatccaaaaacacattgaaatgta 4067
DB 443 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 502
OY 4068 agagattgagatgtaactgtaacgtataatgttgatgaataatggaatgt 4127
```

```
DB 503 AAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 562
OY 4128 atgacattgctgaagcttaataccactacagatgaataatgattcagatgaga 4187
DB 563 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 622
OY 4188 gtttgataataatcattgacatgctcagagcaagaagaataacattgagaaga 4247
DB 623 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 682
OY 4248 ttaacgaaaaaattgctggtacttaaacattcattgatacaactgta 4304
DB 683 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAG 739
```

Search completed: August 25, 2001, 00:08:01
Job time: 6426 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 22:49:15 ; Search time 8259.91 Seconds
(without alignments)
12033.516 Million cell updates/sec

Title: us-09-430-590E-3

Perfect score: 6426
Sequence: 1 tcttggttctgcacattt.....agaagttatattccatca 6426

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 segs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pac1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: gb_pl5:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rtd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rtd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pac:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_r01:*
95: gb_r02:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	DB	ID	Description
1	6426	100.0	6426	12	AF007776	AF007776 Candida a
2	6421.2	99.9	6980	12	AF050215	AF050215 Candida a
3	432	6.7	4872	14	CAU251464	AU251464 Candida a
4	400	6.2	408	12	AF030556	AF030556 Candida a
5	339.2	5.3	1470	12	AF078809	AF078809 Candida a
6	174.6	2.7	1017	53	CNS078UD	AL437951 T7 end of
7	166	2.6	1888	10	AX073190	AX073190 Sequence
8	142.4	2.2	905	53	CNS07ACX	AL436423 T3 end of

9	127.2	2.0	838	53	CNS07ARP	AL436955 T7 end of
C	10	99.4	1.5	80561	12	AC073555 Arabidops
C	11	98.8	1.5	109155	14	ATT20K12 Arabidops
C	12	98	1.5	85961	12	AC079131 Arabidops
C	13	98	1.5	103637	12	AC079604 Arabidops
C	14	97.2	1.5	97146	12	AC007259 Arabidops
C	15	96.4	1.5	4793	13	ATLECR6K Arabidops
C	16	96.4	1.5	13558	14	ATT16L24 Arabidops
C	17	96.4	1.5	91851	14	ATT16L24 Arabidops
C	18	96.2	1.5	1035	53	CNS07AOP Arabidops
C	19	94.6	1.5	82612	14	ATT31B5 Arabidops
C	20	93.2	1.5	82875	12	AC013482 Arabidops
C	21	92	1.4	173565	13	AP003311 Arabidops
C	22	91.4	1.4	16720	14	LEU68072 Arabidops
C	23	89.8	1.4	5846	14	PVA005762 Arabidops
C	24	89.2	1.4	11554	13	AP002460 Arabidops
C	25	89	1.4	152423	12	AC079853 Arabidops
C	26	89	1.4	192174	83	AP003263 Arabidops
C	27	88.8	1.4	145789	60	AP002926 Arabidops
C	28	86.2	1.3	84785	13	AP002064 Arabidops
C	29	86	1.3	157788	83	AP003289 Arabidops
C	30	85.6	1.3	154662	72	AC051632 Arabidops
C	31	85.4	1.3	72520	12	AC006570 Arabidops
C	32	85.4	1.3	77290	12	AC074284 Arabidops
C	33	84.4	1.3	123183	12	AC006841 Arabidops
C	34	84	1.3	110000	84	PFMAL4P1_1 Arabidops
C	35	84	1.3	144322	13	AP002867 Arabidops
C	36	84	1.3	145358	78	AC091086 Arabidops
C	37	84	1.3	147857	13	AP002747 Arabidops
C	38	84	1.3	153477	60	AC006278 Arabidops
C	39	83.8	1.3	13341	14	OLKRLP1 Arabidops
C	40	83.4	1.3	9093	14	AY016208 Arabidops
C	41	83.4	1.3	87286	12	AB026651 Arabidops
C	42	83	1.3	151978	13	AP002069 Arabidops
C	43	82.6	1.3	153776	83	AP003511 Arabidops
C	44	80.8	1.3	155215	83	AP003449 Arabidops
C	45	80	1.2	189349	12	AC066654 Arabidops

ALIGNMENTS

RESULT 1
AF007776 6426 bp DNA PLN 21-NOV-1997
DEFINITION Candida albicans retrotransposon pCal, complete sequence.
ACCESSION AF007776
VERSION AF007776.1 GI:2636718
KEYWORDS

SOURCE

ORGANISM Candida albicans.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1 (bases 1 to 6426)
AUTHORS Matthews,G.D., Goodwin,T.J., Butler,M.I., Berryman,T.A. and

TITLE pCal, a highly unusual Ty1/copia retrotransposon from the

JOURNAL J. Bacteriol. 179 (22), 7118-7128 (1997)

REFERENCE 98037512

AUTHORS Matthews,G.D., Goodwin,T.J.D., Butler,M.I., Berryman,T.A. and

TITLE Direct Submission

JOURNAL Submitted (11-JUN-1997) Department of Biochemistry, University of

FEATURES Location/Qualifiers

Source

1..6426

/organism="Candida albicans"

/db_xref="taxon:5476"

/transposon="retrotransposon pCal"

LTR primer_bind 281..291

CDS
/note="Minus-strand primer-binding site"
398..1372
/note="ORF1"
/codon_start=1
/transl_table=12
/product="Gag protein"
/protein_id="AAC49879.1"
/db_xref="GI:2636720"
/translation="MSSAKNDNEGVESVDQANAKSVDEHAKRFNMLFIRPNL
PKIANGNOKSVDMKEEFKFEVAVADYAEFLDYDNPDKKVEKKVEGYETGNCLO
CIOSIFDRRLMISKLPHKLOKRAKLAIDATYKSDYITTSKIVNENHEL
VVCYNLPYLSQVEKLEETLNTSVNVEVRSPLNGVLYFNHVKSEALSLFLN
IHASYSKNIQADNDTSYVSCSTIAEEMCDHPDARLVDYPSNKEELNLYSLPAPE
KPKGKEENSOSOKNKKSRKRNKKHPSKSDNGEKEKEKTS"

misc_feature

CDS

/note="potentially suppressed stop codon"
41373..6103
/note="ORF2"
/codon_start=1
/transl_table=12
/product="POL protein"
/protein_id="AAC49878.1"
/db_xref="GI:2636719"

/translation="KTGAASINCVANINHCSTTPPVNSHSLNLSLVNMFKLRN
KLYVDYGTATISVNNKDLISNVKATTEVSADATLEADIGDLIRVAGISTLE
NPLYPESEFNLSLKQIEERGFNLIKESVIVNOMVAPPIIISRKNADLVWGP
FSESLCEPDFDGLADMLSNODKDSNENSEVOEHYSRALINSLEYDVL
DVEISRYGEVLEPTGDKNDIYFNHLSHMSIEKILLKQVGLVLTSHESLOKIA
DKVCILSNARKSRHNNHSEKRSRHEHDTIGPERSNNKMYLTSVIDEHTGY
EGITTRKRVKULLIQRKINNNRNDKAYIRSNMAEFPPQSLAFGIFREITIA
YSPELNGLEAVNKLLOOTIRIVTLGLQILKLYVYQSYTHINHTPRSLKGT
PYGCVYQLESEGNRFPFPAIDCVTFESMAIEKNRNVSTGAPSSINGAVIGASDC
FSEYVYLKMKRCDIILSPNVRILRSYEVINSYKMLSTTPMSHIVPAEGLIGRSGA
OYEVRIYVESEYDNTNDVMHMKFSYVOPASFLTGNSENVEYINDDPQITEN
PDQSNPILDTREESHDMYSEVSDNPMPSLHETLPGNPPVSKPPOLGETHRSVGKKS
EPTNNTKAPISIGRDRKRSSTVQVLSHPQCTGASESKISGTDHSGVYVKE
TVSEMDHTSDIPEYSAEDQNPVLSANKNRTEKIDGENTSPFGGDDSVINSNV
EOSNVEDAGNSPIODEVQSGRLINQTLVDVAIVENKISPTINSLSDDHTELA
TDSGNSTESDIOKNEISVINENKTEILIOKHIESILADKRLDEFEYINVEIEN
VINDDIAEANPLDPENNVDYONNESFNNKSRKRYFEKEVNEIACITKSLDT
TDPREAIRVLTNGETKRTEPKREVPIYTKLKRKRSOYSPVYTRSGRFTVIMPKRYLA
VYNKTDYNDPGRVTKSMNALEKFRSKDYEVPIPTGVKPSIMGMVHTEKIDSLGV
KRSRCVHGNRQKELKIDPESVSSVIDVLTLLITIGELGWTIOHLDVESYVLN
ASITSHNPILYFPPKSVPLKRNHMLRSYVGLKSGLEWHTKRVLEDIGETQVL
HNDGFLHEIEEGSVIYGLVYDDILMWSGSKVDNVDOLRHFVEKVGELSNYL
GIEPFTESGYILISOEFLKLLKFDKDDSGKNIPWIPMDKEVKAIIENENPEN
DPEKVNERTLLDPDAKKLYOSGVGLMAATVTRPDISVYVNSIGSKCANPNVDYK
LIVCLARYIKNSGCHIEYKRNMLNTPKSPYIEGSDSPAPGDRKISCTLIYVYK
NLQWATKQIVIAOSSACEMALNTMLKAIETKNNLMDLGEVGIHCHODQAV
IKVLRNNCHPRPIDIQYKFLRQLINKVFISIVYKTDNVDADCTCLSRNFKAF
VEGMIRKLDLEPNQTSIONAIVAE"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

Query Match 100.0% Score 6426; DB 12; Length 6426;
Best Local Similarity 100.0%; Pred No. 0;
Matches 6426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgttggttggtgactatttgggtgcaagaacacgataatgaatgagtgatattatga 60
|||||
Db 1 TGTGTGTTGTGCACATATTGTGTGCAGAAACGATGATGAAATGATGATTATATGA 60
QY 61 gaatgaaattttccatcacacatcaggtgagcagaactaatatattgtgtagt 120

D	b	61	GAATGAAATTTTCCATCACCATAGGTGATGCAGACTAACTTATTGGTAGT	120
Q	y	121	ataaataaggatgataaaccaacatcccaagata tcaagagatagaaggagagt	180
D	b	121	ATTAATAGGGTATGAAATACCAACATCCAGATATTCACGAGTAGAAGGAGAGTT	180
Q	y	181	toaataatattctgtgaataaataacttsgtttaataactataaacaactaagcgt	240
D	b	181	TCATATATATCTGTGTGAATATATACTGTTTAAATTCATATACATACATGACGT	240
Q	y	241	aaacgctcaatctcaaggtaaagaagtttatlcatcaagaattagaatcatagtgat	300
D	b	241	ACACGCTCAATCTCAGGTAAGAAAGATTTATTTCCATCAGATTAGAAGTCATGTGAT	300
Q	y	301	aataattcgcccaaatattagcgttgtataaatltaacgtlcccaagatttgtatttgat	360
D	b	301	AATCATTTTCGCCCAATATGACGTTGATTAATTAATTCAGTCCATGATTTTATTTGAT	360
Q	y	361	gatagtttcgaagttttaagtaagaattltaacaagatgagttccgcaagaatgata	420
D	b	361	GATAGTTTCGAAGTTTCAAGTACAGATATTTCCACAGATGAGTTCCGCCAAATATGATGA	420
Q	y	421	taacgaaggaaaggtlcatggaagtgltgataagctaaatgctatagtaagtgatga	480
D	b	421	TAAAGAAAGGAGGCTCATGAAAGTGTGATCAAGCTAATGCTATTAAGTAAGAGTGATGA	480
Q	y	481	acatacaagctatagttcaataigtctttcataaatlttaagtactaccgaattgac	540
D	b	481	ACATATCAAGGCTGATTCATATGCTTTTCATATAATTTAATGACTTACCTTAAGTTGGC	540
Q	y	541	cgtcggtaaatcaagaagaacgltgataaaltggaatlgaaatlttaatatltccagttgc	600
D	b	541	CGTGGTATCAGAAAGCGGTGATAATGATGAAGAAATTTAATATTTCACGTTCC	600
Q	y	601	ttaccccgatgttttggaaatttttgcttgtaactaatctaaagataaatcaaggttaa	660
D	b	601	TTACCCCGATGTTTGGAAATTTTGGCTGCTGATTAATCTTAAGATAATTTCAAGGTTAA	660
Q	y	661	aaagttgaaggtatttaattltaactgtgtgtgtttacaagaatgttttaagttcaatt	720
D	b	661	AAAGTGAAGGTATTTATTTTACTGTTGGTGTTTTACAATAGTGTTTACATCCATTTT	720
Q	y	721	tgataggttcagatttgatcatgatattctaaagctacaagaacgtgcaaaaggaagcaa	780
D	b	721	TGATAGGTTCAGATTGATGATCATGATTTCTAAGTACCAAGCACTGCAAAAGAGCAAA	780
Q	y	781	cttaatcaaacgtctatataatgcgtttactaaactlaaaggttatatactatcagtaa	840
D	b	781	CTTATATCAAACTCTTATGATGCTGTTACTTAATCTTAAGATATATACATTACTAGTAA	840
Q	y	841	gacttgcttgtaagtttglaaacgttgaaatlgagttagttgttgcataaacttcata	900
D	b	841	GACTTGTGTGAAGTTTATTAACGTTGACATGATGAGTTAGTGTTCCTATTAACCTTCATA	900
Q	y	901	tttgcataagttgaagaagaactlttagaanaataactctacaactctaaagttgttga	960
D	b	901	TTTGTCTCAGGTGGAAGAGAACTTGAGGAATATCTTCAACACTTCAAAAGTTGTGGA	960
Q	y	961	tgaattatgtccgttagtttccaaatctcatagtgtaagcttgtlacttoaalcatgttaa	1020
D	b	961	TGAGTATGTCCGTGATCTTCCAAATGTCATAGTGCAAGCTTGTTACTTCAATCATGTGAA	1020
Q	y	1021	gaaatcagagctttaagtttgttttgaatatlcatgcctcatactactacaagtgat	1080
D	b	1021	GAATTCAGAGCTTTTAAGTTTGTTTTGAATATCTTAAGCTCATCTCTCAAACTGGAT	1080
Q	y	1081	tcaagctgaacatatatcatcagttactcccaagttctctacatatagctgaagaatgtg	1140
D	b	1081	TCAAGCTGACATGATATCATTCAGTACTCCCAATGTCCTTACCATATGCTGGAAGAAATGTG	1140
Q	y	1141	tgatcatccgtgataatgctagattgtgttgacattcccaagcaaaatlgaaactaatct	1200

Db	1141	TGATCATCTCTAATTATGCTAGATTGGTTGACATTTCCAGCACCAATATGAACTTAATCT	1200
Qy	1201	tatltttagtttaccagcaccagagaacccaagaagaaacccagagagaactcacttga	1260
Db	1201	TATGTGTAGTTTACCAGACCAGAGAAACCAAGAAGAAACCAAGAGAAACTCACTGGA	1260
Qy	1261	acaatctcaaaaaagaacctgaaatcaagaaagagaatlaagaaacatccaaatcaga	1320
Db	1261	ACAATCTCAAAAAGAGAAACCTGGAATTCAGAAAGAGAAATTAAGAAACATCCAAATTCAGA	1320
Qy	1321	taacataaaggtbtaaaaaagaagaagaagaacactcactcttgaaatgaataaacag	1380
Db	1321	TAAAGATTAAGGTGAAAAAGAAAAAGAAAAAGAAAAAAGAACTTCACTGGAATGAAAAACAG	1380
Qy	1381	tgctgctctcatctaatctgtgtaatgaaatacaataatgcaagcaaaacccagttccagt	1440
Db	1381	TGCTGCTCTCATTAATTGTAATGTAAGTAATATCAATTAATGAGCAAAACCAACGTTTCCAGT	1440
Qy	1441	agaaatctcatctctctctaaatgcttccttgaaacgttaagaatttaaaagtttaagtt	1500
Db	1441	AGAAATCTCATCTCTCTTAATGCTTTTGAACGTAATGAATTTTAAAAGTTTAAAGTT	1500
Qy	1501	taacaagatcagtgatgatactatggtgcacaaatatctgtgtgtaacaataaagatat	1560
Db	1501	TAAACAGATCTCAATGATGATGATGATGCTGTCGACCAATATCTGTTGTGAACATTAAGATAT	1560
Qy	1561	atttgttgatcttgaagcagcaaaacttgaagtcttctgttgcctgcatggtgtcaactaga	1620
Db	1561	ATTGCTGATCTTAAAGGACCAACAATTAAGTTTCTGTTGCGATGAGTGCTACATTATGA	1620
Qy	1621	agcaagatgtaattgttgatcctaactcaagaagtcggtatgtctgcataagttagaaga	1680
Db	1621	AGCAGATGTATGTGTGATCTAATTAATCAAGATCGGATTTGCTGCATTAGCTTTAGAGA	1680
Qy	1681	tacattgtatllaccagaagaatccctttaaactctgttgagtttgaaacaaattgaagaag	1740
Db	1681	TACATTTGATTTTACCAGAAAGTTCTTAACTCTTGAGAGTTTGAAACAAATTAAGAAACG	1740
Qy	1741	aggaattaaatgcttctaactaaagaacagatgttgatttaaccaaaatgttgctcc	1800
Db	1741	AGGATTTAATTTCTCTTATTACTAAAGATATGATGATTTGATTTTAAACCAAAATGTGGCTCC	1800
Qy	1801	tactattatgctctcaagaagaatgctgtgtaactcttataatggtccctcaatctcagtga	1860
Db	1801	TACTATTATTTGCTTCAAGGAAGAATGCTGTCGTGATTTTATGTGGCTCCATTTACATGA	1860
Qy	1861	agaactctgaagaatgtaatttgaatgaatgtgttggaagaatagtgttccaatgctaa	1920
Db	1861	AGAATCTTTAGAAATGTATTTTGATTTATGATGATTTTGGAAGATATGTTTCCAAATGCTPA	1920
Qy	1921	ccaagatgacaagaataaatacaagatgaaatgaatgtaacgaatatacaagaacagatga	1980
Db	1921	CCAAGATGACAAAGATTAATCAACAGATGATGAAATGTGCAGATTAATCAAGAACATGATTA	1980
Qy	1981	tagttctcagagcatlataaactcttgcgaagggttgatggttttagatglttgaatttc	2040
Db	1981	TAGTTCTGCACATTAATTAATTTCTTTGACGAGGTGTGATGTTTATGATGTGAATTTTC	2040
Qy	2041	cccatatgagttgaacaatgtgtctaccacgtgagaataagaacgatatataacttcca	2100
Db	2041	CCCATATGAGATTTGACAAATTTGCTACCAACATGAGATTAAGAACGATTTTATTAATTTCCA	2100
Qy	2101	tttgaatgcaaaatcaatagtcocatgtgaanaatctgtgtgttacaanaataccaggtct	2160
Db	2101	TTTGATGTCAAAATCATATGTGTCCATTGAGAAATCTTGTTGTATCAAAAATATCCAGGCTCT	2160
Qy	2161	cgtaactcaactccaagaagaagctctccaagaatgttgctgtttlaagatagtctatt	2220
Db	2161	CGTACTTACACTTCAAAAAGAGCTTCCAAAAGATTTGCTGATTTGAAGATATGCTATTT	2220
Qy	2221	atcgaatgccaacagaagaagtcacaatcatcttcaagaagaagaagctctgagaagca	2280
Db	2221	ATCGAATGCCAAACAGAGAGTCACATATCATCTTTGAGAAACAAAAGCTCTGAGAGACGA	2280

QY 2281 tgaagagcttcatgtatatactctcggltccattaggtccgaaataaagaatggtatt 2340
DB 2281 TGAGAGCTTCTATTTGATCTCGGTCCATTAGGTCCGAAATAACAGATGATTTT 2340
QY 2341 aagctcgttataagatgaatacagggltacatgaagaattactactaaacagaa 2400
DB 2341 AAGCTGTTATAGATGAACATACGGGTACATTTGAAGAAATTTATCTAAACAGAAA 2400
QY 2401 ggtaaagatccttcaatcaagaataaagatcgtgaataatcggtttaagataagt 2460
DB 2401 GGTAAAGATCTCTTAATTAACGATTAAGATCTGGAATAATCGTTTAACGATAAGGT 2460
QY 2461 ggcatactcaagaatgataatgctcctgagttcccaacacttcgatttagt 2520
DB 2461 GGCTACTTCAAGAGTATATGCTCTGAGTCCCAACCTTCGTATTTAGCTGAGTT 2520
QY 2521 cggatattgagaggagatacagcgacatacctgctgagcttaatggtctcgagagt 2580
DB 2521 CGGATTTGGAGGAGACTATAGCGCATATCTGAGCTTAAGTCTGAGCTTAAGTCTGAGGAGT 2580
QY 2581 tgttaataaatttgatttacaacagatttacagatcgttggagacttggtccacaat 2640
DB 2581 TGTTAATAATTTGATTTTACACAGATTTACAGGATCTTGTGACACTTGTGACACAAAT 2640
QY 2641 actcaagltgattatataatgaltcaataatctcattacaatgatacacaactccag 2700
DB 2641 ACTCAAGTGTATTTATGATGATTTCAATTTATTAACATGATCAACACACTCCAGC 2700
QY 2701 tctgtactcaagagaaacccctatggttgcattatcaatgaagtgaggaaatt 2760
DB 2701 TCGTTTCACTCAAGGAGAACCCCTTATGATGATGATTAATTAAGTATGAGGAGAAATTT 2760
QY 2761 ctaccggttcccttgcacatcagatgtgtcgttaccattagtaatccatcgaaagaa 2820
DB 2761 CTACCGGTTCTTTCGCACTCGATTTGTGCTTACATTTAGTAATCCATCGAAAAAGAA 2820
QY 2821 ccgltacggagattacatacctaagaagagctcccttcacatcgatcagtggtcgtgattg 2880
DB 2821 CCGTTACGAGATTCATCACTAAAGAGAGCTCTTCATGATCATGGGTCTGTGATTTGG 2880
QY 2881 ctacgagatgattgattgattatctggttgcattcaaaaataatgagtgatattat 2940
DB 2881 CTACGAGTACGATTTGATTTATTAAGTGTGCTTAATAAATAATGAGGTGATATTTAT 2940
QY 2941 ccttagccctaattgtccgtatatactgcaagctatgaggttacttaaccctatccaaa 3000
DB 2941 CCTTAGCCCTATGTCGATATGCGAAGCTATGAGGTATTAATCCATCTCAAAAA 3000
QY 3001 ctatccaactaacctatgtcaacatgttctcctatggtcgaaggtatccaaggagaa 3060
DB 3001 CTATCCACTACCTTACCTATGTCACATGTTGCTTATGCTGAGGATTCACAGGAGAAAGCA 3060
QY 3061 acttgagagctcaofacag 3120
DB 3061 ACTGAGGCTCTAGTACAGAGGATGCGGAAACATATGTGAAAGTGAATAATACATATCAAAA 3120
QY 3121 tgaagtgatgacatgcccagaagatcatatcagttcagcccaagcatcgttacttaac 3180
DB 3121 TGAGTGATGACATGCCCCAAGAGTCAATTCAGTTCAGCCACACATCGTTTACTTTAAC 3180
QY 3181 taagggttaacagttctaagaaatgttataatgataatgatacagttacagatlaacat 3240
DB 3181 TACGGGTAAAGTCTTAACGAAATGTTATTAATGATGATCACTACAGATTAATCAATTTGA 3240
QY 3241 gaatcccgatgatttttctaaacctctcaactaactgaagaatcaacagatatggtatc 3300
DB 3241 GAATCCCGATGATTTTCTTAACCTCTTCAACTAATGAAGAAACACACATATGATATC 3300
QY 3301 cgaagtaaatcgaatgagaaatcctaaacccaggtctccagagtaacacactcgagtaa 3360
DB 3301 CGAAGTAAATCGATGAGATCTTAACCCAGTCTCCACAGAGTAAACACTGGGAGTAA 3360
QY 3361 cccaagaggtatltacatgctgctgcaacaataatcgactataatgatacggatgagat 4500

QY 3361 tccggtgtctaaacctcccaacttgytaccgagatcagtaataaggaatcctaaaga 3420
DB 3361 TCCGGTGTCTAAACCTCCTCAACTGTGATCCGAGAGCTCATATTAAGGAGTCTAAAGA 3420
QY 3421 gccattcaacaacacacaaggagcccttccatccaggaggagagcaataagcct 3480
DB 3421 GCCATTCAAAACCCACACAAAGAGAGGCCCTTCCTCCAGGGAGGGACCTTAACCGCT 3480
QY 3481 ggaactactgtcaggttgaatctatcaacaaaccccaagacttgatcccgcttgcga 3540
DB 3481 GGAATCTACTGCTCAGTGTGAGTACACACCAACCCAGACTGGTACTCCGCTGCGGA 3540
QY 3541 ggaatcaaatgtcagaagagacatcattcgtgtcagctgtttaaagaagaagctc 3600
DB 3541 GGAGTCAAAATGTGAGAGACAGATCATTTGGTGTCCAGCTTGTTAAGAAACAGTCTC 3600
QY 3601 agaagattggcactcttctgactaccagaatacagtgctgaagaatgaacagcaaatcc 3660
DB 3601 AGAAGATTGGCATCTCTTGTGACTACCCAGAAACTATGCTCAAGATGAACAGCAAAATCC 3660
QY 3661 ctcttactggtctaaagaatcgggttaactgaanaaaatagatgagaggagaatatc 3720
DB 3661 CTCGTTACTGCTTATTAAGAAATCGGGTAACTGAATAAATATAGTGGAGAAATATTTC 3720
QY 3721 attccgggggtgataatgatttctgtcgtgatacactcaaatcgttgagcaatcattgt 3780
DB 3721 ATTTCCGGGGGTGATGATTTCTGTGATCACTCAATGATTTGAGCAATCTAATGT 3780
QY 3781 tgaacagagagtgctgttgaagatcgaatcaaacagagaaatttctcaaggagaaat 3840
DB 3781 TGAACAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3840
QY 3841 acttaatgaacaaactgatalatagttgatactgtgtcctaaglttatgagaatgaanaat 3900
DB 3841 ACTTAATGAACAACTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 3900
QY 3901 ctctcctatlaattcaatagatgatacactgaactgtctacagactcgggaaatgatag 3960
DB 3901 CTTCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3960
QY 3961 caatcaacagagatccagatcgaatcgaatgaatgaatgaatgaatgaatgaatgaat 4020
DB 3961 CAATTCACAGATCCGACATTCATTCGAAATGAATATATTCACAGATGATTAATGATAA 4020
QY 4021 aatactgaataatcccaaaaacacatgaagaatccttgcgtgataagagatlgatga 4080
DB 4021 AATACTGAATAATCCAAACAAACATGGAATGATTCCTGCTGATGAAGATTTGATGA 4080
QY 4081 attgaaacgtatlaattgttgaatgaatgagagatgataatgagcgtatgactgtga 4140
DB 4081 ATTTGAAAGGTATTAATGTTGATGAATTAATGATGATGATGATGATGATGATGATGAT 4140
QY 4141 agctaatccactaccagatgaataatgaatgataatgataatgataatgataatgataat 4200
DB 4141 AGCTAATCCACTACAGATGAATAATTAATGATGATGATGATGATGATGATGATGATGAT 4200
QY 4201 tcatagcatgtaacagagcaagaagaatlacacatttgagaagaagtttaacgaaat 4260
DB 4201 TCATAGCATGTACAGCAGCAAGAAAGAAATACACATTTGAGAAAGAAATTAACGAAAT 4260
QY 4261 tgcgtgtactaaacattcactgttatacaactgataccaaagagagaaatcagagtgtaa 4320
DB 4261 TCGTGTACTTAACATTCACTTGTATCACTGATCACTGATCACTGATCACTGATCACT 4320
QY 4321 tactgttgaaccaaagagatcgaacccaagaagaagagagtgctcatcactgtgaat 4380
DB 4321 TACTGTGAACCAAGAGATGAACCCAGAAAGAGAGGCTTATCACTGTAATTT 4380
QY 4381 aaacaaagatcgaaataaagtcacacatglttaacaaagtggtgaagcgttataaa 4440
DB 4381 AAACAAAGATCGCAATTAACATGACATATGTTACAGAAAGTGTGAGACGTTATTA 4440
QY 4441 cccaagaggtatltacatgctgctgcaacaataatcgactataatgatacggatgagat 4500

Db	5521	TAGGATATACAAAATAGCATGGGATATACATTGAGTACAAAAGAACAGATTGAAAT	5580
QY	5581	accaccaaatcatcttgttaltcgaatggttcagtgatgctalttgcacagatttga	5640
Db	5581	ACCACCAAAATCATTTGTATGCAATGTTTCAGTGATGCGTCATTTGCACCAAGATTGGA	5640
QY	5641	tagaaatctattagtggaaacttgattatgttgaaatggaaatttggctgagggcgac	5700
Db	5641	TAGAAATCTATTAGAGCACTTTGATTTATGTGAATGGAATTTGGTCAATGGCGCAC	5700
QY	5701	caaaaacaagaagtcatagcacaagaagctcagaagcttgcgaatttgcgtctaatta	5760
Db	5701	CAAAAACAAGAGTCATAGCACAAGAGCTCAGACAGTTGTGAAATGTTGGCTTAATTA	5760
QY	5761	tacaatgttgaagctatcgcgaataaaaaacatttaalysalttgggttttgaagtga	5820
Db	5761	TACAATGTTGGAAGGATTCGAAATTAATAAACCATTTTAATGGAATTTGGGTTTGAAGTAG	5820
QY	5821	taagaatacttgcatacagaacacgaagcttgattaaagtttggagaataactatgt	5880
Db	5821	TAAGATACATTTCTCATCAAGACACCAAGCTGTGATTAAAGTTTGGAGAAATACATTGG	5880
QY	5881	tcaccacaatcgaaccaatagatatctgtcatalaagtttctacgcacaattatcatgata	5940
Db	5881	TCACCACATCGACCAATAGATATCTGCTATGAATTTCTACGCCAATGATCATGATTA	5940
QY	5941	agattttcaatatcttctgtgaagaacaaatbatatttaagccagatttgatgactaagt	6000
Db	5941	AGTATTTTCAATATCTCTATGTGAACAACAATATTAATTAAGCCGATTTGATATCTAAGG	6000
QY	6001	tctaagtcgtgtcctaattcgaagcatctcgtltgaagtgatgatataaaacggttagactaga	6060
Db	6001	TCTAAGTCGTGTCTAATTCAAAGCATGTGGTTGAGGSTATGATAAACGGTTAGACCTAGA	6060
QY	6061	agatatcaaacacatcgatatcaaaatgcaataacggcagaataagtgattatcatlact	6120
Db	6061	AGATATCAAAACACTGATATCAAAATGCAATTAACGGCAGATAAGTGATTAATTCATTTACT	6120
QY	6121	attatcgtatgtctcaatcgaaggaggtgttggtttgcacattttgtgtcagaactgt	6180
Db	6121	ATTATGTAATGCTCAATTCAGGAGGAGTGTGGTTTGTGCACATTTTGTGTGAGAACTG	6180
QY	6181	atcaatgaaatgatgtgtattatataagaatggaataatlttccatcacacatcaggtagt	6240
Db	6181	ATCAATGAAAAAGATGGTTATTATGAGATGCAAAATTTTTCATTCACACATCAAGTGAT	6240
QY	6241	gacaaactaaactatctgtgtagtataaaagaaggtatgaataccaactccagaa	6300
Db	6241	GACAAACTAAACTATATTGTGTACTATAAATAAGGTAATGAATACCAACTTCCAGAA	6300
QY	6301	tatcaagagatagaaggaggaggttccaatalatalatcttgyaaataaacttctgtct	6360
Db	6301	TATCAACGAGATAGAGGAGGAGGATTTCAATATATATCTTCGATATTAATTAACCTGCTCT	6360
QY	6361	aattcacatacaacacacagtcgttacaagcttcaatctcggfataaagaagtatatatt	6420
Db	6361	AATTACTATATCACACTTAGACGTGTACACGCTCAATCTCAGGTAAAGAAAGTTATATT	6420
QY	6421	ccatca 6426	
Db	6421	CCATCA 6426	
RESULT	2		
AF050215	6980 bp	DNA	PLN 01-JUL-1998
LOCUS			
DEFINITION	Candida albicans Tca2 retrotransposon gag polyprotein (gag) and pol		
ACCESSION	AF050215		
VERSION	AF050215.1	GI:3273716	
KEYWORDS			
SOURCE	Candida albicans.		
ORGANISM	Candida albicans		

Db 747 CGTGGTAATCAGAAAGCGTGTGAATAATGAATGAATTAATTTCCACGTTGC 806
QY 601 ttcccccgaatttttggaaatttttgccttgacttaacttaagaataaacttaagttta 660
Db 807 TTACCCCGAAGTTTGGAAATTTTGTGCTGACTTAATCTTAAGAAATAATTCAGAGTTAA 866
QY 661 aaagtgaaagatttatttacttggttggttttcaaaatgtgttcaagtcattt 720
Db 867 AAAGGTGAAGATTTATTTACTGGTGTGTTTACAAATGTGTTTACAGTCCATTTTT 926
QY 721 tgaatgtcagaatgtgactgaatgttcttaagctcaaaagcacttgcacaaagaa 780
Db 927 TGATAGGTTGAGATTCATGATGATTTCTAAGCTACCAAGCACTTCCAAAGAACCA 986
QY 781 cttaataaagctgccttaataatgtcttacttaataatgaagattatatactta 840
Db 987 CTTAATCAAGCTGCTTATGATGCTGTTAAATCTTAAGATTATACCTTATCTATG 1046
QY 841 gactctgtgaagtttgaataacgttgacataagttagttgttgcctataactccata 900
Db 1047 GACTTGTGCTGAGTTTGTAAAGCTGTGAACATGAGTTAGTGTGTTGCTATTAACCTTCCATA 1106
QY 901 ttctgtcgaagtggaagaagaacttgaagaataactcttaacactcaacgttgcga 960
Db 1107 TTTGCTGCAGCTGGAAGAGAACTTGAAGAAATCTACACACCTTCAACGTTGTGCA 1166
QY 961 tgaatgtcgcgtgactcttcaaatctcaatgactgaagttcaagtttacttaactatgaa 1020
Db 1167 TGAGTATGTCCTGAGTCTTCCAAATCTCATAGTCAAGTCTTGTACTTCAATATGTGAA 1226
QY 1021 gaaatcaagagcttgaagttgttttgaataatcaatgcctcaactactcaagttga 1080
Db 1227 GAAATCAGAGCTTTTAAAGTTGTTTGAATATTCATGACCTCATACTACCAAGTGAT 1286
QY 1081 tcaagctgcaaatgataactcagttactcccaagttgcttaccatactgcgaagaatg 1140
Db 1287 TCAGCTGCAATGATACATCACTACTCCCAAGTTGCTTACCATGCTGAAGAAATGTG 1346
QY 1141 tgaatcactgattatgtctgaatgtgttgaacttcaacgaacaaatgaacttaact 1200
Db 1347 TGATCATCTGATTTATGTATGATTTGGTTGATTCATCCAGCAACAAATATGACTTAATCT 1406
QY 1201 tatgtgtatgttaccagaccagaaagaaacaaagaaacaaagaggaactcactga 1260
Db 1407 TATTTGATGTTTACAGCACACAGAAACCAAGAAACCAAGAGCAACTCATCTGCA 1466
QY 1261 acaatctcaaaagaagaacctgaaatcaagaagaagaataaagaacatccaaatcaga 1320
Db 1467 ACATCTCAAAAGAAAGAACTGAAATCAAGAAAGAAATTAAGAAACATCCAAATCAGA 1526
QY 1321 taacgaataagtgaaagaagaagaagaagaagaagaactcaactlgaatgaanaacag 1380
Db 1527 TAACGATTAAGGTGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1586
QY 1381 tgcgtcttcatattatgtgtgaataactaactatgcagcaaaacaggttccag 1440
Db 1587 TGCCTGCTCTATTAATTTGCTAATGAATATACATAATTCACGAAACCAACCTTCCACT 1646
QY 1441 agaaaatctcatcttcaatgtcttcttgaacgttaacgaattttaaggttgaagtt 1500
Db 1647 AGAAATTTTCATTCCTTAATGCTCTTGAACGTAAATGAATTTTAAAGTTTAAAGTT 1706
QY 1501 taacaaatcagtagtgaataactgcgtgcaacaaatctgtgtgaacaaataagatat 1560
Db 1707 TAACAAATCTAGTGTATGATACTGCTCCACAAATATCTGTTGGAACAAATTAAGATAT 1766
QY 1561 attgcgaatgttaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1620
Db 1767 ATTGCTGAATGTTAAGAGAGCAACATTAAGATTTCTGTTGCTGAGAGGCTGACATTAAGA 1826
QY 1621 agcagaatgtatgtgtactaataatcagagtcggtatgtctcgaatcagttaga 1680
Db 1827 AGCAGATGTTATGTGATCTAATATCAGAGTCGGTATTTGCTGATTAAGCTTAGAGAA 1886

QY 1681 tacatgtattaccagaagaagttccttcaatcttgtagttgaacaaatgaagaag 1740
Db 1887 TACATGTATTTTACCAGAAAGTTCTTAACTTGTGATTTGAACAAATGAAGAACG 1946
QY 1741 agatttaagtcttattactcaagaagaagaagaagaagaagaagaagaagaagaaga 1800
Db 1947 AGGATTTAATGTTCTTATTAAGTAAGATCACTGATTTATTAACCAAAATGTGCTCC 2006
QY 1801 tactatttgcctcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1860
Db 2007 TACTATTAATGCTTCAAGAGAAATGCTGATGATTTATTAATGCTTCAATTCAGTGA 2066
QY 1861 agaatcttgaagaatgtatgtatgaatgaatgtgtgcagatatgtgtccatgtaa 1920
Db 2067 AGAATCTTGAATGTGATTTTGAATTAAGATGTTTGGCAATATGTTGCTCAATGCTAA 2126
QY 1921 ccaagatgaagaagaataatcaatgaatgaatgaatgaatgaatgaatgaatgaatga 1980
Db 2127 CCAAGATGAAGAATTAATCAAGTATCAATGAATGAATGAATGAATGAATGAATGAATGA 2186
QY 1981 tagttcgaagcatataaatacttcttgcgaggttgaatgtttagatgttgaatctc 2040
Db 2187 TAGTCTGAGCATTAATTAATTTCTTGACGAGGTTGATGTTTACATGTTGAATTTTC 2246
QY 2041 cccatagagatgtgaacaatgtctaccactgagagataagaagaatattataatcca 2100
Db 2247 CCAATATGAGATTTGAACAATTTGCTACCAATGAGATTAAGACATATTAATTTTCCA 2306
QY 2101 ttgtatgcaaatcaatcaatgctcaatgaagaatcttgccttcaaaaataccaggtct 2160
Db 2307 TTTGATGCAATATCATATGTCATGAGAAATCTGTTGTTTACAAAATCCAGGCTCT 2366
QY 2161 cgtactcaacttcaaaagaagaagttctcaaaagaatgtcgaatgtgaatgtcact 2220
Db 2367 CGTACTGACACTTCAAAAGAGAGCTTCAAAAGATGCTGATTTGAAGTATGCTAAT 2426
QY 2221 atcgaatgccaaagaagaatcaaatcatcatcacaagaagaagaagaagaagaaga 2280
Db 2427 ATCGAATGCCAAAGAGAGAAAGTCAATCATCATTCAGAAAGAAAGCCTTGAAAGAGCA 2486
QY 2281 tgaagaactcaatgtgactcgtgcacttgaagtcggaataaataaagaatgtgatt 2340
Db 2487 TGAGAGACTTATGATGATCTGCGGCTCATTTAGGTCGCAAAATTAACAGATGCTATT 2546
QY 2341 aacgtctgtatagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2400
Db 2547 AACGCTGTATAGATGACATACGCGGTTACATTGAAGAAATTTACTTAAGACAGAAA 2606
QY 2401 ggtaaagatctcttaattcaacgaataaagaatctggaataatcgttlttaacgaataag 2460
Db 2607 GSTAAGGATGCTTAATTAACAGATTAAGATCTGGAATATGAGTTTAACGATTAAGT 2666
QY 2461 ggcatacttcaagaatgtatgtcctgagttcccaacaaccttgaattagctgaat 2520
Db 2667 GGCATATCTTCAAGAGTGAATGTGCTCCGAGTTCCCAACCTTCTGATTTAGCTGAGTT 2726
QY 2521 cgtatattggaagaagaactatagcgcaatctcgtgaagtttaagttcgcgagag 2580
Db 2727 CGGTATTTGGAGGAGACTATAGCGCATATCTGCTGAGGTTATAGGTTGCTGCGGAGGT 2786
QY 2581 tgttaataatgtatttatacaacagatttaacagatcgttltgaacacttgcgtcacaaat 2640
Db 2787 TGTTAATTAATTTATTAACAGATTTACAGATGCTTGTGACACTTGTGTCACAAAT 2846
QY 2641 actcaagttgtattatlatatgtatcaatattcaatgaatgaatgaatgaatgaatgaatga 2700
Db 2847 ACTCAAGTTGATTTATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2906
QY 2701 tgcgttcaactcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2760
Db 2907 TCGTTCACTCAAGGAGCAAAACCCCTTATGTTGCTGATTTATCAATTAAGTGAAGGAAATTT 2966

QY 2761 ctaccggttccttctgcatcagatgtgtcgttacctttagtaatgcacatcgaaagaa 2820
DB 2967 CTACGGTTCCTTTCGTCATCGATGTCGTTACATTTTGTATGATGCGATCGAAAGAA 3026
QY 2821 ccgttaaggaattacataactaaagagcctctcatcgatcgtatggtctgtatggtg 2880
DB 3027 CCGTTACGAGATTACATCACTAAAGAGCTCTTACATCGATGATGAGTGTGATGAGTGG 3086
QY 2881 ctacgtgagcatgttttagtattatgctgtgtctaaataatgtcggtgtgatatat 2940
DB 3087 CTACGCTAGCCATTCCTTTTACTTATAGCTGTGCTAAAAATATGCGGTGTGATATAT 3146
QY 2941 ccttagccctaatgtcgtatataatgtgaaagctatgaggttataactcctatcctaa 3000
DB 3147 CCTTAGCCCTATGTCCTATATGCGAAGCTATGAGGTATATACCTCTATCTCAAAAA 3206
QY 3001 cttaaccatcactatgtatcacacatgtgtcctatgtctgtgaaggtatccaggaagca 3060
DB 3207 CTTATCCACTACCTACCTATGTCACATGTTCTATGCTGTAAGGTATCCAGGGAAGCA 3266
QY 3061 actggcgctcagtaagagagagcagagacatgtgtgaaggtatgaataatgaatac 3120
DB 3267 ACTGGGCGCTCAGTACAGAGTACGAGGAGACATATGTGGAAGTGAATATGACATACAA 3326
QY 3121 tgaagtatgacatgtccaaagagatcaltcagttccagccagcatcgtttacttaac 3180
DB 3327 TGACGTGATGACATGCCCAAGAGTCAATTCAGTTCCAGCCACATCGTTTACTTTAAC 3386
QY 3181 tacgggtacagttcttaacgaatatgttataatgatatgattccagttacagttaccat 3240
DB 3387 TACGGGTACAGTTCTTACGAATATGTTATATATGATGATCCAGTACAGATTTACCAT 3446
QY 3241 gaatccgaatgttttcttaacccctctcaactacatgaagaatcacagatcgtatc 3300
DB 3447 GAATCCGATGATTTTCTTAACCCCTCTCACTAATGAAGAATCACAGATATGATGATC 3506
QY 3301 cgaagtaaaatcgtatgagatccttaaaaccaggtctccagagctaaacctgtgggt 3360
DB 3507 CGAATTAATATCGATGAGTGAATCCATAACCCAGTCTCCACAGGTATACACTGGGGTAA 3566
QY 3361 lccggtgtctaaacctcctcaactgtgtacagagactcattataggaagtttaaga 3420
DB 3567 TCCGGTGTCTAACTCCTCACTTGTGTACGAGACTTCATATATAGGAAGTCTTAAGA 3626
QY 3421 gtcattatacaaacacacaaagagccctccatccaggggaggaacataagcct 3480
DB 3627 GCCTATTACAAACACACAAAGAGAGCCCTTCATCCAGGGAGGAGCACATAAACGCT 3686
QY 3481 ggaatcactgtcaggtgtgactatacacaaaccagagctgtactcccgcttga 3540
DB 3687 GGAATCTACTGCTCAGTTGATGATACACCAACCCAGACTGATCTCCCGCTTGGA 3746
QY 3541 ggaagcaaatgttcaagaaacagatcaltcgtgtcagagttgttlaaagaacagctc 3600
DB 3747 GGAGTCAAAATTTGTCAGAACAGATTCATTTGCTGTGACCTGTTAAAGAAACAGTCTC 3806
QY 3601 agaagaattgacatctctgtactccagaaacatgtgtcgtgaagaatgaacagcaaatcc 3660
DB 3807 AGAATATTGGCATACTTCTTGACTACCCAGAACTAGTCTGTAAGATGACAGCAAAATCC 3866
QY 3661 ctgttactgtcctaaatgaagatcgtgtactgaaataatagaatgaggaataatcttc 3720
DB 3867 CTCGTTACTGGCTATATTAAGATCGGGTACAGTGAATAAATATATATGAGGAGAAATATTTTC 3926
QY 3721 atttcgggggtgtatgtatctgtcgttgcatacactcaaatgtttagcaatcattat 3780
DB 3927 ATTTCGGGGGTGTATGTATGTCTGTGTGATCACTCAATTTTGAAGCATGTATATGT 3986
QY 3781 tgaacagaggaatgtcgttgaacagttcaaatcgaagagttctcgaagggaggaat 3840
DB 3987 TGAACAGAGAGATGCTGTGTAACAGTCCAAATCAAGACGAAAGTTTCTCAAGAGGAGAAAT 4046
QY 3841 actaatgaacaaactgtatagttgatactgtgtcgttaagttatgtgaatgaaataat 3900
DB 4047 ACTTAATGAACAAACTGATATGATTGATCTGTGCTAAAGTTATATGAATGAATAAAT 4106
QY 3901 ctctcctatataatcattttagatgatcatctgaacttgttacaagctcgggaatgtatg 3960
DB 4107 CTTCTCTATTAAATTCATTTAGATGATCATCTTAACCTGTACAGACTCGGAAATGATAG 4166
QY 3961 caatcaacagaatccgacatcactcaatcgaataatgaatatcacacagttatgaagaa 4020
DB 4167 CAATTACAGAAATCCGACATTCATCGAAATATGAATATCCAGATGATTAATGAGAA 4226
QY 4021 aaatctgaataatccaaacacatgtgaagatccttgcgtgaagaagatgtgatga 4080
DB 4227 AAATACTGAATATATCCAAACACATGAAAGTATCCTTGTGATTAAGATGATGATGA 4286
QY 4081 atttgaacgtataatgtgtgaagaattgaagaatgtgtatgaatgaatgaatgtcga 4140
DB 4287 ATTTGAACGTATATATGTTGATGAATTTGAGATGTGATTAATGACATGATGCTGA 4346
QY 4141 agctaatccactaccagatgaaataatgaatgttccagatgaatgtgaatgttgaataa 4200
DB 4347 AGCTTAATCCACTACAGATGAATAATATGATGTTCAGATGAATGAGATTTGATTAATA 4406
QY 4201 tcaatgcatgttcaagcaagaagaagaataacacatttggagaagaagttaaagaaat 4260
DB 4407 TCAATGACATGTCAAGCAAGAAAGAAATACATTTTGAAGAAAGATTAAAGAAATAT 4466
QY 4261 tgtctgtactaaacattcactgtatcaactgatccaaagagaacatcagatgtttaa 4320
DB 4467 TGTGTGTACTAAATTCATCTTGCATGATCAACATGATCCAAAGAGCATGAGTGTAA 4526
QY 4321 tactgtgtgaaacaaagaagaatcgaacccaagaagaaggtgtcctatcactgtgaat 4380
DB 4527 TACTGTGTGAACCAAGAGAAATGATCCAAAGAAAGAGGCTGCTATACCTGTAATAAT 4586
QY 4381 aaacaaagaatcgaataatgaatgaacatatgttcaagaagaatgttgaaggttataa 4440
DB 4587 AAACAAAGATCCCAATACAGTACAGTATGTTTAAAGAAAGGTGATGACGTTATTA 4646
QY 4441 ccccaagaagttatttcaatgtcgtgtcgtcaacaaatcgcctaaatgtatcgtgagat 4500
DB 4647 CCCCAAGAGTATTTATCATGCGGTGCTCAACAAATCGACTTAATGATCCGGATGAT 4706
QY 4501 aaagtcaatgaatgtcgtgaactagagaatattgatcaaaagatgtttacgaagaatcc 4560
DB 4707 AAAGTCAATGAAGTGAACCTACAGAAATTTAGATCAAAAGATGTTTGAAGAAAGTTC 4766
QY 4561 aattccacacgggtgtgaagcctatcatctatgtgtgtgtgtacatactgaagaatgtatc 4620
DB 4767 AATTCCACCGGTGTGAAGCCTATATCTATGGTGTGGTACATGATGAAATATGATTC 4826
QY 4621 tctcaaaagtgtgttgcgaataatcagttgtgtlccatgtgcaacagaacaaagaa 4680
DB 4827 TCTCAAGGTTGTGTGCGAATAATCAAGTGTGTGCTTCATGCAACAGCAAAAGAAAA 4886
QY 4681 attgagttatgaacctttagtgtgttagtgcacctgttataagtttgyactataagtt 4740
DB 4887 ATTGATTTATGACCTTTTATGTTAGTTAGTTACCTGTTATAGATCTTGTACTATAGAT 4946
QY 4741 attgacaataatagttgtgaatltagaatgtgaacttcaactttagagctcgaatcgc 4800
DB 4947 ATTGACATATATATGTTGTGATTTAGGAATGACAAATTCACATTTTAGAGCTGATCCG 5006
QY 4801 gtaataatgtcctctattactatcactcaatccaaatltatgtcttctcctaaatcagt 4860
DB 5007 GTATCTAAATGCTTATATACATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 5066
QY 4861 accttgaagaacacacatgtgtgttattgaagttctgtctatggtttaaagctc 4920
DB 5067 ACCTTTGAAGAAACCATTTGTTGTTATGAACCTTCTGTATGTTGTTTAAACAGTC 5126
QY 4921 ggtttggaatgtatcacactaaagaagatttgaagaacatgtgtttaccagat 4980

Db 5127 GGGTTTTGAATGGTATCACACTATCAAAAGAGTATTGGAAGACATGTTTATATCAAGT 5186
 Qy 4981 tttaacaatgatggttttttccattgataatgaaagggatcagtaataatttaag 5040
 Db 5187 TTTACCAATGATGGTTTATTTACATTTGAATGAAAGGGATGATATATATTTTAAG 5246
 Qy 5041 ttataltgtatgataattcttattggttggaagttcacaaaagattgataatttgt 5100
 Db 5247 TTTATATGTTGATGATATCTTATGTTGGAATTCACAAAAGTATTGATATTTTGT 5306
 Qy 5101 ggtatcatgagagatcatatttgaagttaaagtgtttggtgaaatacaaatatcttgg 5160
 Db 5307 GGATTCATTTGAGATCATTTTGAAGTTTGAAGTTTGGTGAATTCATAATTTATCTGG 5366
 Qy 5161 tatgatttcgttaaaacccaatcigtatatttatcttccagaanaaaatttctcaaga 5220
 Db 5367 TATTGATTTTCGTAAACCCGAACTCTGTTATTTATCTCAAGAAATTTCTCAAGAA 5426
 Qy 5221 attactaaagatttcaaaactagatgactatataatggaanaaaacataccctgattccga 5280
 Db 5427 ATTACTTAAGATTTCAAACTAGATGATCATTTGGAATAAAACATTCCTGATTCGAA 5486
 Qy 5281 tgaacaatatagaagaagttgcaataattcgtgtaaaacgttaataccagagaatgttga 5340
 Db 5487 TGACAAATATGAAAGGTTGCAATTAATTCGTGAAACGTTAATCCAGAAATGATTTTGA 5546
 Qy 5341 aaaggtccgaatgagacattgcttgacctga tgcataaaactataccctgatttgaatc 5400
 Db 5547 AAGGTTCCGAATGAGACATTTGCTTACCCCTGATGCTTAAAAAATTTACCAAGTGTGT 5606
 Qy 5401 tggctgctttatagggtcgcaaaaacacgltccagatatcgttcgtatgtaattc 5460
 Db 5607 TGGCCGCTTTTATGCGCTGCCCAAAACACACCTCCAGATTAATCGGTGATGAATTC 5666
 Qy 5461 gttgggttccaaatccgaatccaaatgctgaatgttggaattgataattatttct 5520
 Db 5667 GTTGGGTTCTAAATTCGCAATTCGCAATTCGATTAATGGAATTTATTTATTTCT 5726
 Qy 5521 taagtalataaaaatagatggagatalacatttgatcaaaaagaacagattgatat 5580
 Db 5727 TAGGTATATCAAAAATAGATGGATGATTCATTCATTCGAAAGAAACAGATTTGAATAT 5786
 Qy 5581 accacaaaatacatctgtatagatggttcagtgatgcgtcaatttgcacagatttga 5640
 Db 5787 ACCACCAAAATCATTTGTTATCGAAAGTTGAGTGCATTCATTCACCGAGATTTGA 5846
 Qy 5641 tagaaatcatatagtggaacttgaatttatgtgaatgaaatttgggtgcaatggcgac 5700
 Db 5847 TAGAAATCTATTAGTGGAACTTTGATTAATGATGGAATTTGCTCAATGGCGCAC 5906
 Qy 5701 caaaaaacaacggtcatagcaaaaagctcagcagcttgaatttgcgtctaata 5760
 Db 5907 CAAAAACAACGCGTCATAGCAACAAGCTCAGAGCTGTGTAATGTTGGCTTAATTA 5966
 Qy 5761 taaaagtgtgaagctatcgaataaaaaacatlaattgatttgggtttggaatgag 5820
 Db 5967 TACAAGTTGGAAGCTATCGAAATTAACAAATTAAGGATTTGGGTTTGAAGTAGG 6026
 Qy 5821 taagatacatgtcatcaagaacaacaaagcttgatlaaagtttggagaataactatgt 5880
 Db 6027 TAGGATACATTTGTCATCAAGACACCAACAGCTGATTAAGTTTGAAGAAATTAACATTTG 6086
 Qy 5881 tcaaccacacatgacaatagatatctgctalaagtttccagcccaatgatalcaatgatala 5940
 Db 6087 TCACCCACATGACCAATAGATATCTGCTAATAGTTTACGCCAATTCATCATGATTA 6146
 Qy 5941 agtatattcaatatctatagtgaagacaatgaataatcgcgattgtatgaactagt 6000
 Db 6147 AGTATTTTAAATATCTATAGTGAAGCAATATTAATTCGCCGATTTGATGACTTAAGTG 6206
 Qy 6001 tctaaatcgtgtaaatcaaaagcattgtaggtatgatatgaataaagcttgaactaga 6060
 Db 6207 TCTAATCTGCTAAATTTGAAGACATTCCTTAAGGATATGATTAAGGTTTGAACCTTAA 6266

Qy 6061 agataatcaaacactgatacaaaaatgcaataacggcagaataagtgattatattact 6120
 Db 6267 AGATTAATCAAAACACTGATTAACAAAATGCAATTAACGCGACAGAAATTAAGTATTAATTA 6326
 Qy 6121 attatcgttaatgctcaatcaggagggtgttggttgcacatttgggtgcacaaactg 6180
 Db 6327 ATTATCGTAATGCTCAATCAAGGAGGTGTGTTGTTGTCACATTAATTTGTGCACAAACTG 6386
 Qy 6181 atcaatgaanaatgaltgtattatgagaatggaatatttccatccacacatcagtgat 6240
 Db 6387 ATCAATGAATAATGATGGTTTATTAATGAGAAATGAAATTTTCCATCAATCAAGTAT 6446
 Qy 6241 gacagaactaaactatattgtgtgataaataagggatgaataatcacaatccaga 6300
 Db 6447 GACCAACTAAACTATATTTGTTAGTAAATAAGGATTAATAATACCAACATCCAGAA 6506
 Qy 6301 latcaacgagataagaggaggtttcaatatatacttgtaataaactcgttct 6360
 Db 6507 TATCAACGAGATAGAGAGGAGAGAGTTCAATATATATCTTGTGAATTAATTAATCTGTTCT 6566
 Qy 6361 aattcaatatacaacaactagacgtgtgatacgcgtcaatctcaggtlaaagaagttatatt 6420
 Db 6567 AATTCACTATACACAACTAGACGTTACACGCTCAATCTCAGTAAGAAAGTTTATATT 6626
 Qy 6421 ccaatca 6426
 Db 6627 CCATCA 6632

RESULT 3
 CAL251464/c
 LOCUS
 DEFINITION
 CANDIDA ALBICANS ENGL GENE FOR ENDO-1,3-BETA-GLUCANASE AND PCAL
 RETROTRANSPON.
 ACCESSION
 AJ251464
 VERSION
 AJ251464.1 GI:6562333
 KEYWORDS
 ENDO-1,3-BETA-GLUCANASE; ENGL GENE; RETROTRANSPON.
 SOURCE
 CANDIDA ALBICANS
 ORGANISM
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Mitoportic Saccharomycetales; Candida.
 ESTEBAN, P.F., Garcia, R., Rios, I., Vazquez De Aldana, C.R. and del
 Rey, F.
 Cloning and characterization of the
 endo-1,3-beta-glucanase-encoding gene CaENG1 in the yeast Candida
 albicans
 Unpublished
 2 (bases 1 to 4872)
 REFERENCE
 del Rey, F.
 Direct Submission
 Submitted (07-DEC-1999) Del Rey F., Departamento de Microbiologia y
 Genetica, Universidad de Salamanca, Campus Miguel de Unamuno,
 Salamanca, 37007, SPAIN
 FEATURES
 source
 1..4872
 Location/Qualifiers
 /organism="Candida albicans"
 /db_xref="taxon:5476"
 1..4440
 /organism="Candida albicans"
 /strain="ATCC64385"
 /db_xref="taxon:5476"
 711..4148
 /gene="eng1"
 711..4148
 /gene="eng1"
 /codon_start=1
 /transl_table=12
 /evidence=experimental
 /product="endo-1,3-beta-glucanase"
 /protein_id="CA62579.1"
 /db_xref="GI:6562334"

LOCUS	AF030556	408 bp	DNA	PLN	01-JUL-1998
DEFINITION	Candida albicans retrotransposon pcal ITR, and GAG gene, partial cds.				
ACCESSION	AF030556				
VERSION	AF030556.1	GI:3273502			
KEYWORDS					
SOURCE					
ORGANISM	Candida albicans.				
REFERENCE	Eukaryote, Fungi; Ascomycota; Saccharomycotina; Saccharomycetales; mitosporic Saccharomycetales; Candida.				
AUTHORS	1 (bases 1 to 408)				
TITLE	Goodwin,T.J.D. and Poulter,R.T.M.				
JOURNAL	Temperature- and strain-dependent expression of the pcal retrotransposon of Candida albicans				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 408)				
TITLE	Goodwin,T.J.D. and Poulter,R.T.M.				
JOURNAL	Direct Submissions				
FEATURES	Submitted (21-OCT-1997) Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin, New Zealand				
SOURCE	Location/Qualifiers				
	1..408				

BASE COUNT	143 a	63 c	75 g	127 t
ORIGIN				

Query Match	6.23:	Score	400:	DB 12:	Length	408:
Best Local Similarity	98.88:	Pred.	No. 8.8e-55:			
Matches	403:	Conservative	0:	Mismatches	5:	Indels
						Gaps
0y	1	ttgttgatttggaacattttgtgtgcgaacatgcatcaatgaataatgatgtattatga	60			
Db	1	TGTTGGTTTGTGCACATATTTTGTGTGCAGAAACATCATCAATGAATATGTTATATGA	60			
0y	61	gaatggaaaatttttccatcacacatcaagtygtatgacgaactaaactatattgttagt	120			
Db	61	GAATGAAATTTTTCATCCATCCATCATGAGTACAGACAGAACTAAACTATATGTGTAGT	120			
0y	121	ataataaagggtatgaatatcccaacatcccgaatatcaagaatatagaaggagaggtt	180			
Db	121	ATAAATAAGGGTATGAATAATACCACATCCCGAATATTCAGAGATATGAAGAGGGGAGGTT	180			
0y	181	tcaatatatctctgttgataataacttcgcttcctaatccactatacaacaactagaagtyt	240			
Db	181	TCAATATATATCTGTGCATATATTAATCTTCGTTCTTAATTCACATATACCAACTAGACGTGT	240			
0y	241	aaacgctcaactcgaagtaaagaagaagtattatattccatcagattagaagtcgaatgtagt	300			
Db	241	ACACGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAATATAGAAAGTCAAAAGCAT	300			
0y	301	aatcatcttcgtcccaaatatgacgttgtatataaatcgaagtcctcagaatttgtattatgatt	360			
Db	301	AACCAATTCGCCCCAAATATAGCGTTGTATTAATTCAGCTCCCAATTTGTATTTTGATT	360			
0y	361	gatagtttcgaagtttgaaggtacagaatttcacaagaatgagttccgcg	408			
Db	361					

Db 361 GATAGTTTGGAGCTTGAAGTACAGAAATTCACAGATGATCCGC 408

RESULT 5

LOCUS AF078809 1470 bp DNA PLN 03-AUG-1998

DEFINITION Candida albicans Tca4 retrotransposon reverse transcriptase (pol)

ACCESSION AF078809

VERSION AF078809.1 GI:3377678

KEYWORDS

SOURCE Candida albicans.

ORGANISM Candida albicans

REFERENCE 1 (bases 1 to 1470)

AUTHORS Goodwin, T.J.D.

TITLE Tca4, a Ty1-copia retrotransposon from Candida albicans

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1470)

AUTHORS Goodwin, T.J.D.

TITLE Direct Submission

JOURNAL Submitted (16-JUL-1998) Department of Biochemistry, University of Otago, Cumberland Street, Dunedin, New Zealand

FEATURES

source 1. 1470

/organism="Candida albicans"

/strain="SC5314"

/db_xref="taxon:5476"

gene <1..>1470

/gene="pol"

/note="Tca4 retrotransposon"

/rpt_family="Ty-1-copia"

/rpt_type="dispersed"

<1..>1470

/gene="pol"

/codon_start=2

/transl_table=12

/product="reverse transcriptase"

/protein_id="AAC28090.1"

/db_xref="GI:3377679"

/translation="GRSVHPKRYLINALIVKKIDYNOKEMROSMEEIEKFRANQVTV
EKTPIKVVPLTMVHTKIDNLKNNHNSKRVGVNMMNBPDPHAISSPVADLT
SIRLSAIAVENNLHOLDIASALINSLDGRVIEVRPGEFVGVKXSRHKSIV
YGLRGSANHWYSHKENVLEANGKOTLNDGIPKNNENGVLYVYVDVEMIAN
EKIIEFVAMLETFOLOYFGEATEYIGTOBRKPDCTLOIIPLEKLVATFNIDS
YKQIPIPKDINNVKQIRKSSQINDPVKLEPKQLINSVKTYIDYEBNEEDHK
ESPQEPISAKGILQSAVSLMASNTPDLSAFVNGAKCAHPDVEDMWRMY
CLRTYKMDRFLKRYKGRLLNKSDFIECFSDASFDPIDRKSITGTSIFVNGNLV
AMATKKOKIITHSSSAECMLALNVTIILKAFDLRLNTIEDL"

BASE COUNT 522 a 241 c 299 g 408 t

ORIGIN

Query Match 5.3%; Score 339.2; DB 12; Length 1470;

Best Local Similarity 54.4%; Pred. No. 4.3e-45;

Matches 790; Conservative 0; Mismatches 578; Indels 84; Gaps 2;

QY 4423 tggtagaagcgtttaaaccacgaagagattacatcgctgctcaacaatcgacta 4482

DB 1 TGGAGATCAGTGCATCCCTAAAGATATTTAATGTCATTGTTAAGAAATAGATTA 60

QY 4483 taatgacccggatgataaagatgaatgctgaactagagaatttagatcaaaaga 4542

DB 61 CAATCAAAAAGATGCGCTCAAGATATGGAAGCAAGAAATCGAAAAATTAAGGCTAACCA 120

QY 4543 tggtagaagagatccaatccacccgggtgtgaagctatataatcggttggtaca 4602

DB 121 AGTTTACCCGTTGAAAAACCAAGACCTGTCCCATTTGAAGAACATGTGGCTACA 180

QY 4603 tactgagaatgtatctctcaaggtgtgtgtcgagaatcacgtgtgtgtgtcaccag 4662

DB 181 TACTTACAAAACATGACCTCAAAATATCATATTACAAAACCCGTCGTTAATGCG 240

QY 4663 caacagacaaaaggaataattgattatgacccctttagttagttacccgttataga 4722

DB 241 AAACATATATGTCGAAATTCGTGATTTTGATCCCATCCATCTCTCCCGGTAGTAGA 300

QY 4723 tcttgtagcttaagatattgacataatagtgtgtgaatttagaataatcaaca 4782

DB 301 TCTCACAAGTATACGACATTTATCTGCGCATAGCTGTTGAAATATATGTTGTTACACA 360

QY 4783 tttagacgtcagtcggtgctataatgctccttactcattcaatccaatccattatgt 4842

DB 361 AFTGACATTCGCCCTCAGCTTATTTGAACGCCAGTTTGAGAGATGGAAGATCTTTGT 420

QY 4843 cttccctcctaataatcagtaaccttgaagaaaacacattgtgtgtttagaagcttctgt 4902

DB 421 GAGACACCCGCGTGTGTTTGAAGTTAAACCTGCGTATAGTTGGCGTTTACACAAGCTGT 480

QY 4903 ctatggtttaaagacgtcgggttggaaigtatcacacatacgaagaagataggaga 4962

DB 481 GTACGCTTTAGCGAGAGTCCCATTAATGTTACTCACATTTTAAGATGTTGGAGGC 540

QY 4963 catggttttaccagtttaccacaaatgattgttattccacattgaataatgaagaagg 5022

DB 541 AAATGCTTTAAACAAACAACTACAAATGATGGCATTTTGGAAAAATTAATGAAATG 600

QY 5023 atcagtaataatagttatagttatagtgtatgataatcttaatgtgtggaagttcaaaa 5082

DB 601 AGATGTATTAATATGATGTATGTATGTGAGATGATGTTTATGATGATCTAATATGAAAA 660

QY 5083 agttatgataatcttggatgataatgagaagatcatttgaagttgaagttgtttgtga 5142

DB 661 GATTATTAAAGATTTTGTGCTATGCTCGAAGACATTTTTCAGTTAAATATTTTGGCGA 720

QY 5143 aataacaaatattctgatttgaatttcgttaaaacccgaatcgtgtattatattatca 5202

DB 721 GGCTACGAGATATCTAGGAATACAGTTTAGAAGACACCGATGATGATACGTTAGACCA 780

QY 5203 agaaaatttctcaagaataatctaaggaatttcaactgagaagatccatagtggaagaaa 5262

DB 781 AATACCATTTCTTGGAAAAATTTGTCGTACATTCAATATTCAGATGTTATGAAAAAGA 840

QY 5263 cataccctggattccgaatgacaaatattg----- 5291

DB 841 CATTCCTATTATTCACAAAAGATATTATGTGTGTCACACATTTGAGAAAGTCAAGTCCAGAT 900

QY 5292 aaaaagttgcataatctcgtgaaac----- 5317

DB 901 AAATGATTTCTGTAAAGTTGGAGAAACCTCAAAAGTTAATCAATTCGCTGTCAAAAACCA 960

QY 5318 -----gttaatccagaagaatgatttgaagaagttccgaatgagac 5358

DB 961 GTACATTGATGATTATACAGAAATATGAAAGACATGATTTTAAGAAATCACCAAGCTGA 1020

QY 5359 atgtctgaccctgattgctlaaaaaactataccaaagtggtgtgtgacctgtttatgggc 5418

DB 1021 ACCCTTAAGCCCAAAAGAAATTAATTAATACCAATCCGAGTGGAGCTGTTGTTATGAGGC 1080

QY 5419 tggcacaacaacacgctcagatataatcgtcgtcagtgatgaattcgttgggttcataatcgc 5478

DB 1081 TTGATGATGACCTTAGGCCAGACTTATAGCTTTTATGTTAAATACCTCGAGCAAAAGTGC 1140

QY 5479 aaatccaaatgtccatattatgagaatgatttattgtctttaggtatatacaaaatag 5538

DB 1141 TCACCCAGATGTCGACAGACGGAAGATATGATGATATGCTTACGGATATTTTAAGAAAGA 1200

QY 5539 catggaatcacatctgagtagaacaagaatgaaatgaaatataccacaaatcatttgt 5598

DB 1201 TATGACCTTCAAACTTAAATATACAAAGAGAGACATCAACAAAGTCAAAAGATTTTAT 1260

QY 5599 tatcgaatgtttagaagatgcattttagaagaagattgagaataatctatagtggt 5658

DB 1261 TATGATATGTTTCTACATGCTCTCATTTGCCCCAGATTTGAGACAAAGTCAATCACCG 1320

BAC clone F114 is from Arabidopsis thaliana chromosome 1. The orientation of the sequence is from SP6 to T7 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including GenScan+ (Chris Burge, <http://CCR-081.mt.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), Glimmer4 (a variant of Glimmer, see Mihela Pertea, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/repeatmasker.html>).

FEATURES

source

1. 80361

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="1"

/clone="F114"

mRNA join(762..801,1066..1183,1260..1333,1447..1570,1805..1987,2071..2312)

/gene="F114.1"

/note="similar to 40S ribosomal protein S7 homolog GI:5532505 from (Brassica oleracea)"

CDS join(1109..1183,1260..1333,1447..1570,1805..1987,2071..2190)

/gene="F114.1"

/codon_start=1

/product="40S ribosomal protein S7 homolog, putative"

/protein_id="AAG60128.1"

/db_xref="GI:1259781"

/translation="MESAQHKIKRKGVELSEDOVAQAFDLENTNOELKSELKDLVNSAVQDISGKRAIVNVPRLKAKRIHYRLVELEKESGKDVILAIRRIYRPPKSGSAKRPNRKLTLSVEALIDVLPALVIGKRTIRYRLDQTKIMKVELDPKERNNTEKVAFAFSAVKKLKGKDVFEFPITTEA"

mRNA complement(join(<2677..2964,3055..3303,3412..3508,4553..4771,4862..5230,6284..6492,6561..>6815))

/gene="F114.2"

/note="similar to 40S ribosomal protein S7 homolog (Arabidopsis thaliana)"

CDS complement(join(2677..2964,3055..3303,3412..3508,4553..4771,4862..5230,6284..6492,6561..>6815))

/gene="F114.2"

/codon_start=1

/product="terpene cyclase, putative"

/protein_id="AAG60130.1"

/db_xref="GI:1259781"

/translation="MEATRVGKLGSKLISHSQNAFPAFKLSRPFITSPFGNAHLPLKATTHPLADEENRERKNGSPSEHQFSAHVDISEMDALEREIALPKYRDMLSSESSKRIIFLYLVSLGLAHFDEDEIKESLDEGLQKIEEMASSEDLEKFDNGKEFCEIADAKGILSLYEAAHMGTTTDDILDEALSETTLVMEASGACKINTSRRIKALDPOHKMEIIVAKMYIOFYEEEDCDTKLKPAKLNKRFLOIHTYDLEKTLISKWKKDOFKSKLPPYFEDRLVECHFASTGCPKRYARARFETSKITVQITDITDQRYASIGVEESLADTERPDHMDGIDPTLKSVKFNVTDFEERKCKRSRLINLQVAKWKAGHLPSDEYLDVAGLELAISFTAGILMEVNCPEAEVEMWSRDLVREVIKRLINDLFEYEDMRGVYTNINCKKOYGTAEAIKRLQIYVADGEEKMMNBEFLKPIVNPQVPRVIVLIDTLRAANVSEKDEDFTRGHEHLKNCITSIYVPL"

repeat_region complement(4261..4281)

repeat_region complement(5501..5529)

repeat_region complement(5593..5617)

repeat_region complement(5735..5755)

repeat_region complement(6117..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

repeat_region complement(6171..6171)

Oy 5033 tattagattatattgtagatattcttattggtgaagttcaaaaagtattgat 5092
 Db 15881 ATCGCATCTTATATGATGACTGATTCACGGGTAACTCCAGCATGTCGAA 15822
 Oy 5093 aatttgtagatattgtagatattcttattggtgaagttggtgaatataat 5152
 Db 15821 GAATTCAGAAAGAGATACGAGAGTTCGAGATGACGACATTCGATTCCTTAC 15762
 Oy 5153 tatttgtagatattgtagatattcttattggtgaagttggtgaatataat 5212
 Db 15761 TATCTCGAATTCGATTAACAGCAAGCAATGATTCATACGACAGAGCTAT 15702
 Oy 5213 ctcaagaattacttaagaattacttaagaattacttaagaattacttaagaatt 5254
 Db 15701 GCTAAGAGGACTTAAAGAGTCAAGATGATGACTCAAT 15660

RESULT 13

AC079604/C

LOCUS AC079604 103637 bp DNA 19-JAN-2001
 DEFINITION Arabidopsis thaliana chromosome 1 BAC T15M6 genomic sequence,
 complete sequence.

VERSION AC079604
 KEYWORDS AC079604.5 GI:12321249
 HTG.

SOURCE
 ORGANISM
 Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 AUTHORS

1 (bases 1 to 103637)
 Liu,X., Kaul,S., Town,C.D., Bentio,M., Creasy,T.H., Haas,B.J.,
 Wu,D., Maiti,R., Rensing,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
 Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
 Arabidopsis thaliana chromosome 1 BAC T15M6 genomic sequence

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BAC clone T15M6 is from Arabidopsis thaliana chromosome 1
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

Genes were identified by a combination of several methods: Gene
 prediction programs including Genscan+ (Chris Burge,
<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/Genemark/>), Glimmer4 (a variant
 of Glimmer3, see <http://www.tigr.org/softlab/glimmer4/glimmer4.html>), and
 GeneSplicer (Michael Perle and Steven Salzberg, contact
perle@tigr.org), searches of the complete sequence against a
 peptide database and the plant EST database at TIGR
<http://www.tigr.org/tdb/cgi.shtml>). Annotated genes are named to
 indicate the level of evidence for their annotation. Genes with
 similarity to other proteins are named after the database hits.
 Genes without significant peptide similarity but with EST
 similarity are named as unknown proteins. Genes without protein
 or EST similarity, that are predicted by more than two gene
 prediction programs over most of their length are annotated as
 hypothetical proteins. Genes encoding tRNAs are predicted by
 tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 Simple repeats are identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

Location/Qualifiers

source
 1. 103637
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="1"
 /map="7"
 /clone="T15M6"
 complement(1.. 663)
 /gene="T15M6.26"
 /note="pseudogene, hypothetical protein"
 /pseudo
 complement(1003..1088)
 /rpt_family="AT_rich"
 complement(1091..2595)
 /gene="T15M6.25"
 /note="pseudogene, hypothetical protein"
 /pseudo
 complement(<3359..>4729)
 /gene="T15M6.24"
 complement(3359..4729)
 /gene="T15M6.24"
 complement(3537..4729)
 /gene="T15M6.24"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AAG50699.1"
 /db_xref="GI:12321255"
 /translation="MNTBASKQTTAEGSANDPEPQILSPRSLEKOKKMWISVLC
 IFVLVGLDVLMLNLFYVDNRSDSDDDIYRTGTLQALVONAFPLLPFFIFP
 SPKQNETNRELSRFLILYISGLVVAHAKLFLAGKLYANGFVTLISATQIL
 TAFPAIIRKFTFRLIILISGLIVFGSPERGSPDEEYFSIOAWLTPAASV
 APALSLCEFLCEKYLKTRKGRKVRMYEIQICVFATVVCVGLFASGK
 ELQDSHREKGEYVYVLSIGLISMOWMAVGLVLYSGVGVGDVHCTSPVLA
 LVVLAFPMDPDSFPRGTIATVVALGSIYTLHRRKKKWEYLQTENNIDV"
 complement(6149..6169)
 /rpt_family="AT_rich"
 complement(6713..6740)
 /rpt_family="AT_rich"
 complement(7047..7147)
 /rpt_family="AT_rich"
 complement(8659..8680)
 /rpt_family="AT_rich"
 complement(10552..10573)
 /rpt_family="AT_rich"
 complement(10742..10991)
 /gene="T15M6.23"
 /note="pseudogene, putative ribosomal protein"
 /pseudo
 complement(11760..11838)
 /rpt_family="AT_rich"
 complement(12070..12708)
 /gene="T15M6.4"
 /note="pseudogene, putative non-LTR retroelement reverse
 transcriptase"
 /pseudo
 complement(18837..19207,19528..19647,19786..19886,
 20080..20291,20407..20787,20902..20989,21228..21288,
 21891..22125,22758..22835,22952..>23614)
 /gene="T15M6.22"
 complement(18837..23614)
 /gene="T15M6.22"
 /note="contains similarity to cell proliferation related
 protein GI:9931486 from (Mus musculus)"
 complement(18837..19207,19528..19647,19786..19886,
 20080..20291,20407..20787,20902..20989,21228..21288,
 21891..22125,22758..22835,22952..23614)
 /gene="T15M6.22"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG50696.1"
 /db_xref="GI:12321252"
 /translation="MKRRGRHKGRKSKTINQGNLNESTENAENQTSRHSSEAVEC
 ENSELESKMEVADPSPIAGSVDLADNIAKSVARVKKTKTSKADPEPNTLLDDIDKS

[illegible]

Query Match	1.5%	Score 98	DB 12	Length 103637
Best Local Similarity	49.2%	Pred 1.3e-06		
Matches 375	Conservative	0	Mismatches 375	Indels 12
			Gaps	4
QY 4496	tgagaataagtcgaatgctgaactagagaaatttagatcataaagaatglttaacgaaga	4555		
Db 88790	tgagaaatccatctggatgtagaagatcmaatcmaatncamaaagatgacactgggaatt	88731		
OY 4556	gttccaatcccacccggtgtgaagccatatactatlygggttgggtacactgagaatt	4615		
Db 88730	acttgcacttccaaatggacacaaagcccaattggcgtgaagtggtgtatmaagcaaa	88671		
OY 4616	gattctccaaagtggtttctc---gaatcaagtttgtttccatggtcaacaagaca	4672		
Db 88670	aactcttaaaagaaagatggaaagatncacaaagcattggtttgcaaaaggttatgca	88611		
OY 4673	aaggaanaatttgattatgacctttagttagtlltcaactgttataagctgtact	4732		
Db 88610	agagcccggaattgactatgac-----gaggtattttgctccggttgcgtatgaanaag	88557		
OY 4733	ataagattatgcaataataggttgtgaatagaatgacaatccaacttaagacgtc	4792		
Db 88556	gtttagactatfctactctactagggcgtcaaaacaaagtgaagatcactcaaatggatg	88497		
OY 4793	gagtcgcgcatcaatcccttacttactcaatcaatcaacttctgtccctcc	4852		
Db 88496	aagtcggcctcttcttaattggaatctttgaagaagaattttacatttgacacacacagc	88437		
OY 4853	aaatcagttaccttgaagaanaaaccaattgttgttattgaagcttccgtatggta	4912		
Db 88436	tacatgactcaaaagtgacaaagacaaagcttggacacttcaaaaagcgctttatggat	88377		
OY 4913	aaacagcgcggttggaaagtgatacactatcaaaagatgtggaagacatggtt	4972		
Db 88376	aaacagcccccaagacttggat-actcgatttgacaaagattttcaaggaagaaagttt	88318		
OY 4973	actcaagtttacaacatgatgttatttcaacttgataatgaagaggtacagtaata	5032		
Db 88317	catcagaatg-ccatattgacgatgcgactcttatcmaaatccaaaagaaagatatttg	88260		
OY 5033	tatttaagttatactgtatgatatcttctatglttgaagttccaaaaaagtattgat	5092		
Db 88259	atccgaatgcttatattgactgattgattttacagggttaacattccaaagcattgtcgaa	88200		
OY 5093	aatttggatacaattggagagatcattttagaagttgaagtttggtaataataat	5152		
Db 88139	gaatttcagaaagaaagatgacgaagagtttcgaatgacgacatttgattgattcttgc	88140		
OY 5153	tatttggatgaattcgtaaaacccaatctgttatacttatactcacaanaatt	5212		
Db 88139	tatttcggattgaaagtaaaacaaagacacattgaaattttcatttactcaaaagcgtat	88080		
OY 5213	ctcaagaacttactaagatttcaaaatagatgactcaat	5254		
Db 88079	gcttaagaggtacttaagaagttcaagatgattgactcaaat	88038		
RESULT 14				
AC007259/c				
LOCUS	97146 bp	DNA	PLN	17-AUG-1999
DEFINITION	Arabidopsis thaliana chromosome I BAC T28P6 genomic sequence,			
ACCESSION	AC007259			
VERSION	AC007259.4	GI:5306232		
KEYWORDS	HTG.			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
AUTHORS	Ekayawa; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta			
	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids III			
	Brassicales; Brassicaceae; Arabidopsis.			
	1 (bases 1 to 97146)			
	Federspiel,N.A., Palm,C.J.,Conway,A.B., Conn,L., Hansen,N.F.,			

Altefi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V. S., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

Unpublished

2 (bases 1 to 97146)

Federfspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altefi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

Direct Submission

Submitted (08-Apr-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

3 (bases 1 to 97146)

Federfspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altefi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

Direct Submission

Submitted (01-Jul-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

4 (bases 1 to 97146)

Federfspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altefi, H., Araujo, R., Huizar, L., Rowley, D., Brooks, S., Buehler, E., Chao, Q., Dunn, P., Gonzalez, A., Khan, S., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vyotskaia, V., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.

Direct Submission

Submitted (17-Aug-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

On Jul 1, 1999 this sequence version replaced gi:5103851. Bases 90095-97146 of clone T28P6 overlap with bases 1-7052 of 'TAU', BAC clone T19D16, gb|ATU95973.

e-mail for correspondence: arabidopsis.stanford.edu

Genes with similarity to proteins in the databases are described as 'putative', 'like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Gail (Informatics Group, Oak Ridge National Laboratory), http://compbio.ornl.gov/section/index.html, GENSCAN (Chris Burge, http://genome.stanford.edu/chris/GENSCAN.html), Fexa (V Solovayev & A. Salamov, Sanger Centre, http://genome.sanger.ac.uk/), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).

Location/Qualifiers

1. 97146

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="1"

/clone="T28P6"

864. .3964

/gene="T28P6.1"

/note="RK51"

join(864. .2139,2289. .2408,2497. .2670,2798. .2917, 3106. .3343,3422. .3566,3650. .3964)

/gene="T28P6.1"

/note="identical to cDNA receptor-like protein kinase clone, g114008007"

/codon_start=1

/product="receptor-like protein kinase"

/protein_id="A04998.1"

/db_xref="GI:5734723"

/translation="MKVFLVFFFLPFCISVDTIMRQSLRDGEVILSAGKRFAPG FFLSDSELRYGIMVQISQGITVWVANDHPINDISGKVSNNRNLVSVDNMT

ELIWTNSVDSMLEPTLVATLSDGLNVLFPDVTGKSPFESDHPDIDPELPMRLGFT RKGLDRSLTSMKSHGDPSPGLILMERGEPOLLYGVPMWMMGKTRHSGV PEPDITVIRNNSSVNNEDNEPFGYVGTADSVITRMVNTGTMRTVATARKRNDF MSYPRQCDNYAICGPNNGVCDSPSSSTFECTCLPGREPFPHMELRDSGGCTKKR ASICSEKDGMDLARTYLNQDGYIIRVKEELARNNRNLGSKRRVLLI,LSLAAV ALICLWHGMDARTYLNQDGYIIRVKEELARNNRNLGSKRRVLLI,LSLAAV MLTIVILFCVRERRKSNRRSSANPAPDFPDESFEDDOKARNPDLPLDNTI VAATNNSSONKAGRVKPYDSSGEVEVERKGTNGRGOAGADIKVAASKSHEE ORLELPMKREMEVIRGIARTIYHODSRLRIIRHDKASNLILDSMPTKSDRGA RIRGDMGSCSRVYGTGGYAAPEYAMGORSIKSDYSRVLVLEITTKKSNAPH EESSNLVGHWDLMENGEATETIDNIMDETDEREVMKICITGLCVOENSDSDVM SSVIMLGHNAITLVPKHPAFISARRRGENGACLGOTGISVNDVTFSDIQGR

complement(4664. .7708)

/gene="T28P6.2"

complement(join(4664. .4972,5054. .5204,5275. .5427, 5661. .5821,5908. .6101,6189. .6296,6391. .7708))

/gene="T28P6.2"

/note="Very similar to receptor-like protein kinase"

/codon_start=1

/protein_id="A04998.1"

/db_xref="GI:5734724"

/translation="MVSVYIRRRFVLLIACCLISRLRCGEEDITFSPKIDSES ETLCKSGIERCEFPFVNSTRLRVGTGWYKPIQTVYVWANNDSPTNDISYISI YODGNLAVTDGRNRRLVSTNSVPAVPAATVQMDSGMLQDRNNGETILWESFKH PYDSFMRMTLGTGDRGTGNLKTWSHSDSTGNYTAGLAPTEPELLMKNVPT WRSGPNNGQVETLPMNDSILFDGLNLSNDQGTISMSYANDSFMYHNLDPBGIIY OKWSTSMRTWIRGVKPYTDDDAYRCRCSFCHAGENPCKCYKGNKNTWNG GNMNSCKMRAPLDCGRORNSVGGGGAGDGLKOKKVPKIVISERSASQVPCV KCLNCSCTAYADRGICGMALNSGDLYDMQSPGLSGDITFIRAHSEILTHSLAWIA APYIGMLIAVCVILLACKRYKRRPAKDRSAELMFKRMEALITDNSASQILAKE LPLEFQVYATSTDSFSLRNKLQGGGQVYVYGRKLPEGDELAVKRLSRKSGGLELM NEVVYISKLQHRNLVLLCCIEGERIDSLRIIRHDKASNLILDSMPTKSDRGA LARIFRANDEANTRVVGTGYGMSPEYMEGFSEKSPVSLGVIPLFETISGRNDS SHKEENNLNLIYAMKLMDGEALSLADPAVDKCEKTEICVHIGLICVOEVANDR PMSVYIMLTTEENSLADPKOPAFIVRGADEALSSDSSQKSVINDVSLTAVGR"

complement(9352. .10839)

/gene="T28P6.3"

complement(join(9352. .10047,10150. .10617,10729. .10839))

/gene="T28P6.3"

/note="Hypothetical protein"

/codon_start=1

/protein_id="A04999.1"

/db_xref="GI:5734725"

/translation="MDPLASVLEKYSKFTSSQDLYVRHGFENPSRLNPIDILKRL QRTFSDMLKREOREKVERIISQKLSGGFQSTHVRREVLYALAILMGNTDE ESNFGNMGKVRGGLSRVETSLRETRDLAEVLAVAGNSGRDRLGLGLSLAKY FYKADINDMFSALIPVGAHFDIASVYSQGLSEVSELGPLLSOHNGSAIGL TPKASNTITSLAOSMSNVREGLDAPNCPFPPOVCHILRSKLILGCHOVLTSSNLHSVGAITVYPSLIRPRTNPEPSPPLMRRNNHVSSTAIKLSIMDEST KICGRTIENSREKQVYKSVSTTNDPEDEVGMSVGVVDSNRHDOFOVEBYSILKFN IDDRFLSPCLVYHNDSGRTIGMLKSHWSL"

14210. .17181

/gene="T28P6.4"

join(14210. .14332,14521. .14578,14694. .14959,15053. .15169, 15263. .15323,15394. .15557,15641. .15735,15822. .15871, 15958. .15998,16078. .16251,16342. .16377,16463. .16530, 16620. .16680,16774. .17181)

/gene="T28P6.4"

/note="Highly similar to MLO proteins; Highly similar to barley MLO (control element of plant pathogen resistance) multi-gene family in arabidopsis"

/codon_start=1

/protein_id="A04999.1"

/db_xref="GI:5734726"

/translation="MADQKERTLETSTWAAVYCVILFTSTVLEHSITKIGTWK KKHQALPALKYVAELMLLGLFISLLITQPTPSNICISQKASTHPPSABEAK KTKGDAGKRDGDDGKRGRLLELASVTHRSLATKGYDKCAEKVAFVSYGI HOLHILFVLAVVAVYCVIYAFGKIKMRWKSWEETKTIQYQSNDRPFRFARD TSGFRHLNFMKSTRVTLVLCYCFRPGFSGYKVVYLLRLGFIIMAHNAPDNESFDR RYIORSLEKDKETVVEISPVIMFAVLEFLNLSGYSMLPELPIVLVILLYGTL RYITIKRIGRIODEKGDVYAGVAPGDDLPMPGKPRILILHILVLTPTMAGLAFPA MSTYEPNLNCHHESTADVITRLVYGAVOILCSYVTLPLTALYTMGSKKPKPYEND

Db	Accession	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	JOURNAL	TITLE	AUTHORS	COMMENT	FEATURES	Source
Oy	4891	gaaacgtctctatcgttgaatggttaaacacagtcggttgtaatggtatcacactataaag	4950												
Dp	33906	AAAAAGGTGCTTTATGATTAAMAACAAGCCCCCAAGACCTTGGAAT-ACTCGAATTGACA	33848												
Oy	4951	agatctgaagaacatctggttttactcaagtttaccacatgataagtttattcacatgta	5010												
Dp	33847	AGTATTTCACAGACAGAAAGATTTCATCAAGTGT--CCATATGACATGCACCTTATATCA	33790												
Oy	5011	atataagaaggaatcagaataataataggttataatglttgataatcttatagtttg	5070												
Dp	33789	AATTCAAAAGAGCATATATGATGCGATGCGTATATATGATGATGATTCACGCG	33730												
Oy	5071	aagtcacaaaaagatctgataatcttgatgataatggaagtcgaatgtaagttga	5130												
Dp	33729	TACCAATCCAAACATATTCGAGATTCGAAGAAGATGACGAAGAGATTGGAATGAC	33670												
Oy	5131	agttgttggaataatcaaatatcttgatgataatcttgataatctgtaaacgaatc	5190												
Dp	33669	GGACATWGTGATGATGCTGTACTATCTCGAATTAAGATTAACAAAGAACATGGAAT	33610												
Oy	5191	tatttactctcaaaaaaatttctcaagaataatcttaagattcaactcaatagatc	5250												
Dp	33609	ATTCTATCTACAGAGAGCTATGCTAAGAGAGTACTTAAGATTCAGATGATGACATC	33550												
Oy	5251	atat 5254													
Dp	33549	AAAT 33546													
RESULT	15														
ATLECRECK	LOCUS	ATLECRECK 4793 bp DNA	PLN	04-FEB-1999											
DEFINITION	A.thaliana lectin receptor kinase gene.														
ACCESSION	Y08010														
VERSION	Y08010.1 GI:1769896														
KEYWORDS	Art1, copia-like element; lectin receptor kinase; transposable element.														
SOURCE	thale cress.														
ORGANISM	Arabidopsis thaliana														
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 4793)														
AUTHORS	Herve,C., Serres,J., Dabos,P. and Lescure,B.														
TITLE	Nucleotide sequence of DNA art1 (accession no. Y08010), a new Arabidopsis copia-like transposable element, disrupts a lectin receptor kinase gene in the columbia ecotype (PGR97-059)														
JOURNAL	Plant Physiol. 113, 1464-1464 (1997)														
AUTHORS	2 (bases 1 to 4793)														
COMMENT	Herve,C.														
FEATURES	Direct Submission														
Source	Submitted (13-SEP-1996) C. Herve, INRA-CNRS, LBMRPM, Chemin De Borde Rouge Bp27, Castanet-Tolosan Cedex, 31326, FRANCE														
	Overlaps with Y08020.														
	Location/Qualifiers														
	1. 4793														
	/organism="Arabidopsis thaliana"														
	/sub_species="Columbia"														

